AD-A049 350 AIR FORCE GEOPHYSICS LAB HANSCOM AFB MASS F/G 9/4

A NEW AUTOMATIC PROCESSING TECHNIQUE FOR SATELLITE IMAGERY ANAL--ETC(U)

AUG 77 R S HAWKINS

UNCLASSIFIED AFGL-TR-77-0174 NL

| OF |

ADA049350

END

DATE FILMED

3-78

DDC

ADA049350

AD No.
DDC FILE COPY

(2)

AFGL-TR-77-0174
AIR FORCE SURVEYS IN GEOPHYSICS, NO. 371

A New Automatic Processing Technique for Satellite Imagery Analysis

R.S. HAWKINS

3 August 1977

Approved for public release; distribution unlimited.

METEOROLOGY DIVISION PROJECT 8628
AIR FORCE GEOPHYSICS LABORATORY
HANSCOM AFB, MASSACHUSETTS 01731

AIR FORCE SYSTEMS COMMAND, USAF

This report has been reviewed by the ESD Information Office (OI) and is releasable to the National Technical Information Service (NTIS).

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

John Howard

Chief Scientist

Qualified requestors may obtain additional copies from the Defense Documentation Center. All others should apply to the National Technical Information Service.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (14) AFGL-TR-77-0174, AFGL-AFSG-371	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) A NEW AUTOMATIC PROCESSING TECHNIQUE FOR SATELLITE IMAGERY ANALYSIS.		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER AFSG No. 371
7. AUTHOR(s) (10) R. S. Hawkins		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Geophysics Laboratory (LYU) Hanscom AFB, Massachusetts 01731		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62101F (16) 86281202 (17) 12
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Geophysics Laboratory (LYU) Hanscom AFB, Massachusetts 01731		12. REPORT DATE (11) 3 August 1977
		13. NUMBER OF PAGES 69 (12) 69p.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (9) Air Force surveys in Geophysics,		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DDC

RECEIVED FEB 2 1978 B

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Image analysis
Satellite image analysis
Redundancy reduction
Data compression

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A new approach to the analysis of satellite imagery is presented. The central part of this approach is an algorithm which compresses information stored in the ordinary six or eight bits per picture element into only one bit. The quality of this compression is demonstrated by examples of its application to high resolution visual imagery. Both visual inspection and rms difference criterion are used for this evaluation. There are four objectives of this report which are: (1) to review the status of processing techniques which remove redundant information, (2) to show the need for redundance reduction in the → next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20. (CONT)

processing of satellite images, ~~(3)~~ to present the development of an algorithm for reducing it, and ~~(4)~~ to show results obtained by application of the algorithm to visual imagery. Also, comments are made on needed developments of the technique and its potential application to problems of analysis of satellite imagery data.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Preface

I would like to thank Lt Colonel Donald Varley of the Meteorology Laboratory, AFGL for reading the manuscript. His many suggestions greatly improved its readability. Thanks to Mr. Donald Cozzens of Regis College under contract to AFGL for his assistance in computer programming for this study and for his operation of the McIDAS system. Also, to Mr. Greg Hunolt of NEPRF, Monterey, California, thanks for supplying digitized tapes of DMSP data.

PRECEDING PAGE NOT FILMED
BLANK

Contents

Illustrations

A New Automatic Processing Technique for Satellite Imagery Analysis

1. INTRODUCTION

The research described in this report began as an effort to develop a technique to concentrate high resolution satellite imagery information into one bit per picture element for use with special purpose imagery channels. The subject arose with respect to processing problems foreseen by adding certain imagery channels to the Defense Meteorological Satellite Program (DMSP) satellite, and in particular to the proposed 1.6 μm snow/cloud imagery channel.*

It was noted in early discussions that sensors could easily provide high data rates, but the processing of another six bit image channel at high spacial resolution would possibly be more than could be justified. The vast amount of data associated with DMSP imagery channels and their related transmission, storage, and processing expense place serious limitations on the number of channels that an operational analysis system, and in particular that at the Air Force Global Weather Central (AFGWC), can handle. In fact, the large amount of data places limitations on all aspects of its use.

(Received for publication 2 August 1977)

*Snow in the 1.6 μm range is a very poor reflector and appears black in imagery at that wave length. This property provides a means for distinguishing snow from clouds. A first generation sensor is currently under development to fly on future DMSP satellites.

In discussions concerning an operational snow/cloud channel, the possibilities of quantizing the channel at some low brightness level to a one bit image were viewed as a possible solution. This would provide high spacial resolution of snow fields which was the primary purpose for the channel. On the other hand, information in the range of brightness of clouds would be lost along with information on cirrus clouds which is another high priority parameter.

As observed at that time, techniques are needed to reduce to manageable volume data rates of very high resolution imagery while retaining much of the fine scale information. A part of that problem which is considered here is that of increasing the concentration of information in a one bit presentation over that obtained by quantizing or, for that matter, any other way. The fact that most imagery, and satellite imagery is no exception, has a low information rate per bit prompted a search for a technique that would transfer those small pieces of information to one bit. The method found to do this is the subject of this paper.

The preceding comments represent a brief outline of the requirement for the work reported on here. Soon after the route for a numerical solution was discovered, it was realized that the technique to be described is of much more general importance to image processing than that of reducing the amount of data. It is this more general standpoint that is to be presented.

The next two sections of this report provide an introduction to automatic processing of satellite imagery that relates to image redundancy (Section 2) and preprocessing (Section 3). These sections provide an orientation important for appreciating the remaining sections.

The goal in the processing of satellite imagery should be to maximize the use of automatic processing equipment and minimize the use of the human analyst. The validity and necessity of this view has become increasingly apparent in recent years as a result of increased data flow from more satellites, more channels, and higher resolutions in both space and time. This requirement for automatic processing restricts the area of search for solutions as much, if not more so, than the types of imagery involved.

As far as known by the author the algorithm that forms the core of this paper has not been previously formulated nor explored. A summary of the early stage concepts of development leading to the construction of the algorithm will be given after some details of a finite interpretation of imagery are presented. The algorithm will then be applied to high resolution visual imagery and the results discussed. These results show that a high degree of image integrity can be maintained while reducing the number of coding bits per picture element from six to one. An objective evaluation of the algorithm is also given.

The results reported here shed new light on problems of image redundancy and information extraction.

2. SATELLITE IMAGE REDUNDANCY REDUCTION

Satellite pictures, as well as most other types of imagery, contain large amounts of redundant information because of statistical interdependencies of picture samples. Recognition of this is not new. Glaser[1], before the first meteorological satellite was launched, gave a very clear description of digital data redundancy to be expected in the imagery and the problems involved in decreasing it. Although an early start was made with this problem, there have been but few and relatively minor advances. One explanation for this may be related to the fact that the early coding techniques were generally very demanding computationally and did not provide sufficient reductions to make them worthwhile. Advancing techniques make it necessary to reconsider conclusions and impressions obtained only a few years ago. Is there hope for development of redundancy reduction techniques? What can they contribute to the satellite imagery data problem? These are important questions that must be encountered and assessed.

In this report redundancy means "multiple statement of image information" in the same sense as it is used in Information Theory. The information specifically referred to here is in digital form. Redundancies of meteorological information–overstatements in an interpretive sense–are not included in this discussion. Such considerations are treated more directly in the context of recognition and extraction, which is further downstream in the process than the present discussion.

This concept of superfluous data can be expressed vividly in a colloquial manner. Suppose there are two types of gasoline, A and B. Gasoline B is a concentrated form. Two gallons of B give essentially the same results (mileage and otherwise) as ten gallons of A. But gasoline B is more expensive because it requires more processing, P, than A. To use B also requires a processing mechanism, p, in the car. There are other less easily described and evaluated factors entering into the situation. Our problem: which gasoline should be used? Should we continue to use A or make arrangements to use B? Obtaining the best answer requires a thorough analysis of the whole system.

This simple analogy illustrates the image analysis problem as related to redundancy. It is believed that a 5:1 ratio for redundancy is reasonable and may even be conservative. Others may argue, however, that essentially the same results of using "all" digital data and using only (perhaps) one-fifth of them cannot be demonstrated. But if by results are meant the overall output of the current system then there is a good chance that reduced data can be as effective. The argument then seems to fall back on processing agents P and p, where P represents reduction or coding, and p represents restoring or decoding before use.

1. Glaser, Arnold H. (1957) Meteorological Utilization of Images of the Earth's Surface Transmitted from a Satellite Vehicle, Harvard University, Blue Hill Observatory, 145 pp.

In the example above, the original, A, and processed, B, versions were taken to be equivalent except for differences in "concentration". This illustrates the idea of redundancy. Now suppose the processed version, B, is in a form which does not require a restoring process, p, before it is used. This points up a special type of redundancy reduction. In this case the process P does two things. It removes redundancies from the data while at the same time it retains information in a useful format. The process can also be considered as a special type of data compression. Apparently this process has not earned sufficient status to have a name of its own. One will not be needed for it here, although the concept will appear several times in what follows. Later, in discussing the first stage of processing of satellite imagery from the standpoint of redundancy and data form, it will be referred to simply as "pre-processing".

A theory of imagery does not exist as yet which would provide some guidance in selecting approaches, aid in following through on the development of techniques and suggest quantitative ways for evaluating results. It is necessary to pick and choose, try different approaches, and to rely heavily on experiments rather than following a set of sound rules.

2.1 Survey of Methods

A cursory view of redundancy reduction techniques relevant to the satellite imagery analysis problem provides specific examples of the kinds of situations encountered.

With both data compression and extraction of information in mind Marggraf[2] developed a technique for encoding elemental features from satellite images. Basic patterns were coded in blocks of data and overall intensity values were permitted to vary from block to block. This approach has apparently not been developed further or applied in the field. It has some desirable features. Data reductions range from 5 to 1 for 3×3 elemental areas to 18 to 1 for 6×6 elemental areas. The technique places imagery data in a one bit plane which is very convenient for developing relationships for comparing with other data. However, the varying brighness level aspect of the process places some restrictions here. And from an operational point of view, encoding and decoding computation times are relatively small. The main question that has not yet been answered relates to the efficiency of the method for retaining image information.

The considerable redundancy in satellite imagery becomes evident when we consider that a picture often consists of large areas of the same brightness level. Examples are the great expanses of open water and large areas of bright clouds. For such cases it would probably be sufficient to transmit data only on the

2. Marggraf, W.A. (1967) Information Content, Elemental Feature Extraction and Coding of Meteorological Satellite Television Data, General Dynamics Report No. GD/C-ERR-AN-1053, unpaged.

boundaries of the homogeneous areas along with information on their content. This would, excluding complications introduced by regions of great variability, result in a smaller number of bits of data as compared to the standard sample by sample method of encoding. However, there are complications here too; all but the simplest data fields are too irregular for this approach.

Statistical dependencies between areas of constant brightness level can be broken up by a technique called "run length coding". Rather than specifying data sample-by-sample, this method encodes the length of similar brightness levels in a scan. When long lengths of similar levels occur much reduction in data amount results. This basic form is probably not applicable to satellite imagery where great variations occur over small areas as in the video data. Coding of distances between successive contours as pointed out by Glaser[1] has serious limitations of a similar nature.

Coding techniques like those discussed above, which rely on structural simplicity of the data, cannot be expected to be very useful when applied to most types of satellite imagery. The images, by and large, are too complicated. Even though there may be large areas with very little variation, there frequently are other areas of tremendous variation and irregularity. To be effective, a technique must account for this in an effective way.

The classical method of redundancy reduction is apparently not in use for processing satellite imagery. In its basic form this approach produces very limited data reductions. This so-called optimum coding technique, in which the code word length is chosen according to the symbol probability, gives a reduced bit rate as compared to fixed code-word length coding. Only limited experimental results of this approach are available (Marggraf)[2] and there are apparently no results for spacial resolutions higher than three nmi. A number of different models were examined by Kutz et al[3] for compressing satellite imagery with radio transmission problems in mind. Their results for this probablistic approach are nevertheless of interest. The sequential structure of these methods is very attractive from the standpoint of applications to real-time processing; but, on the other hand, the encoded data, in compressed form, have to be decoded before they can be used. And this defeats the purpose, from the standpoint of analysis, of removing redundancies.

The direct use of the classical method does not appear promising; however, there are indirect approaches in use in digital TV transmission that cannot be discounted. The methods are used in connection with DPCM (differential pulse code modulation) where successive differences of one form or another are encoded.

---

3. Kutz, R. L., Sciulli, J. A., and Stampfl, R. A. (1968) Adaptive data compression for video signals, Advances in Communication System, Vol. 3, edited by A. V. Balakrishnan, Academic Press, New York, pp 29-66.

This is the most commonly used technique for compressing digital TV signals (Häberle et al;[4] Mussman,[5] Kummerow[6]).

Classical transformation methods are often used in imagery processing to contend with the problem of statistical dependencies between picture elements. The most widely used are those of Fourier, Hadamard, and Loève-Karhunen. Pratt[7] compared the three transforms from the point of view of data compression. On the basis of mean-square-error of reproduction, he found the Loève-Karhunen was superior, followed by the Fourier and then the Hadamard. In terms of ease of implementation it is the reverse order. Another important paper on this subject is by Habibi.[8] Both papers (7 and 8) are reproduced in a book by Davisson and Gray[9] which contains reprints of some of the most significant papers in the new and rapidly expanding field of data compression. The major drawback in these transforms for satellite image compression is computational complexity and time required for execution. There may also be questions as to the desirability of the form of the data for purposes of information extraction.

Linear filtering techniques are being developed for data compression requirements. It has been shown that linear transformations are equivalent to linear filtering operations which are usually much easier to realize on a real-time basis (Häberle et al;[4] Mussmann[5]). These studies are specifically related to digital TV transmission. They apparently have not yet been applied to the analysis of satellite imagery.

2.2 Information in Compressed Data Forms

Techniques for removing image redundancy can be separated into two distinct groups: those that are meaningful in the compressed form and those that are not. The former is of limited value for real-time processing and will not concern us here. The latter group does not have a specific name but encompasses a number of generic terms such as information extraction, feature extraction, pattern recognition, classification, etc. These reduced data states contain information in one or both of two forms: spacial and spectral. The classical terms are time and frequency. Or, as we are concerned with distance instead of time - space and frequency.

These two attributes of the data are not quantitatively defined but depend qualitatively on the nature of the information that is most accessible. If distances and sizes of features are easily obtained, then it has spacial information. If on the other hand general statements such as "widespread brightness with few irregularities" or "mostly covered with many very small features" are valid descriptions, then the

---

(Because of the large number of references cited above, they will not be listed here. See Reference Page 67 for References 4 through 9.)

data are said to have spectral information. Of course if both types of information are accessible then the data have both properties.

For real-time image processing of a general nature both types of information should be present in a form that can be rapidly used. That is, techniques for reducing redundancy in imagery should change the data to a form in which both spacial and spectral information is readily available. From a computational standpoint, the more accessible the information the better.

Imagery in its basic form is spacial by nature and its equivalent spectral form is obtained through one of a number of mathematical transformations. These two data forms, called "domains", are traditionally treated as remote from one another, however, they have been shown to be equivalent in terms of information. That is, one is a redundant statement of the other.

Haralick and Shanmugen[10] discuss the importance of these two aspects of imagery and show how they can be combined for classification purposes. They place considerable emphasis on textural features and provide a set of sixteen basic descriptors for their determination. This approach bypasses some of the problems discussed above in that information extraction is approached without giving any specific attention to problems of redundancy. When statistical relationships are found between the extracted features and the parameters sought, this is fine; but when they are not found, there remains much uncertainty. In this case it may mean that (1) other extractors are needed, or (2) another approach is needed, or (3) there is no relationship between the images and the sought after physical information. Methods that put image properties together arbitrarily require a rather definite statement of the types of features expected to be important. When this is available, they can give good results.

2.3 Spacial and Spectral Domains

Various techniques have been devised to circumvent problems encountered in the initial stages of image processing. Where specific end results can be seen in advance this is undoubtedly the best approach. When the process must be flexible for many potential applications, these initial stages are very important. In this respect, questions arise in connection with the image processing technique.

Are the spacial and spectral domains separate and distinct by necessity or it it possible to have both readily available for processing purposes without undue duplication? Must we maintain one data set in two forms separated by a significant amount of computation to have immediate access to the information of the data set? With certain qualifications and limitations it can be shown that the two domains need

10. Haralick, R. M., and Shanmugan, K. W. (1974) Combined spectral and spacial processing of ERTS imagery data, Remote Sensing of the Environment 3:3-13.

not be remote from each other, and that both spacial and spectral information can be readily available for analysis purposes without mutual redundancies.

Methods designed to reduce redundancy and put data in a favorable form for extracting information will be referred to as pre-processing techniques. Some general aspects of such techniques have been discussed. The next section is more specific in this regard.

2.4 Definitions of Image and of Frequency

Before proceeding further the concepts of image and frequency must be better defined for purposes of this report.

By image is meant a digital representation, and in particular a two dimensional array of positive numbers representing areal brightness values. The problem of obtaining these numbers, referencing them geographically and correcting them for various effects, although important, will not be considered here. Also the question of what differences would result if measurements had been made at some small displacement from that given will not concern us. We assume all normalizing corrections have been applied and that information such as sun elevation, nadir angle, etc. that may be required for subsequent interpretation is retained.

Although the word frequency is used sparingly, the concept as used here must be clearly understood for it is implicit in most of the arguments of this report. It is the statistical-probability meaning of frequency that is used here which refers to a number of entities over a time or space interval rather than the engineering-analytic concept based on a sine wave representation. A good glossary of other common terms as well as special technical terms used in image processing is given by Haralick.[11] Also an excellent reference in this regard is the new book by Rosenfeld and Kak.[12]

3. PRE-PROCESSING SATELLITE IMAGERY

Why pre-process satellite imagery? Why not design specific extraction techniques operating on the basic data? These questions are not answered simply but there are some simple observations that may be made with which to weigh the importance and consequences of the two approaches. Pre-processing may be used to simplify the data form without significantly changing its content. One image is a very complicated statement of information, and several types of images (for example,

11. Haralick, R. M. (1973) Glossary and Index to Remotely Sensed Image Pattern Recognition Concepts, Pattern Recognition, Vol. 5, Pergamon Press, pp 391-403.

12. Rosenfeld, Azriel, and Kak, A. C. (1976) Digital Picture Processing, Academic Press, New York.

visual, IR, microwave, 1.6 μm) greatly increase this complexity. Consequently, considerable computational effort is required to evaluate these images. On the other hand, any efficient pre-processing procedure requires a considerable amount of computational effort that is not directed toward a specific solution. Techniques that start from basic data are, in general, not hampered by interface and compatibility problems that can result when starting from pre-processed data.

The best case in favor of pre-processing can be made where a number of specific techniques require essentially the same processing in early stages. When there are no reasons to consult the original data, the process can save much in the way of computer effort and make possible more extensive evaluations of the imagery. Thus, the degree of generality plays an important part in whether or not to incorporate pre-processing into an automatic imagery analysis system.

It is essential that the first stage of image processing retain a spacial format in order to be sufficiently general for extensive applications. In addition, both spacial and spectral information on a local scale should be either explicit or readily available from the processed form.

Any approach taken will quickly show that these idealistic criteria are not independent. Any decision about one or two physical aspects of pre-processing either implicitly or explicitly puts considerable confines on other aspects of the problem. Further, the requirement for retaining local information places a limit on the minimum size that data can be reduced to in terms of digits. That is, one bit per picture element.

3.1 Idealized Configuration for Information Extraction

Figure 1 is a very basic and extremely general statement on an analysis system designed to obtain meteorologically useful information by automatic processing techniques applied to satellite imagery. It represents the process of information extraction. For an approach or reduction technique to be viable, it must make a computationally feasible connection from image to result. The most effective connections between the left side and the right side of the figure must be sought. There is no guide indicating how to do this or even how to find out when a connection does not exist. In the Figure, image properties mean any digital statement of one or more of the input images. In practice, they would have a more specific nature. Textural measures, gradient measures, areal means and variances of image brightness are some categories of what may be referred to as image properties. An image property is a specific statement in digital form that describes an image. It may be local or global or both, spacial or spectral or both; it may be very general or very specific.

Generating image properties is not a problem; however, obtaining a connection between them and meteorological parameters has been successful on only the most

rudimentary levels. Perhaps, (1) the image properties aren't adequate or (2) the meteorological parameters are not stated satisfactorily or (3) relationships exist but haven't been detected, or (4) relationships do not exist.

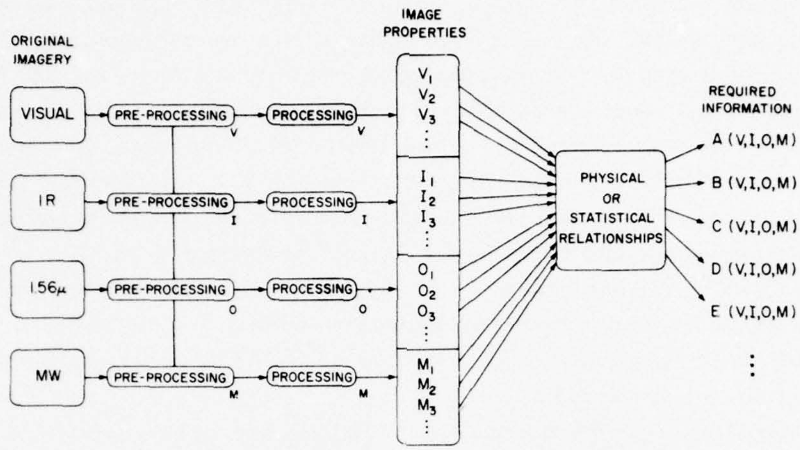


Figure 1. Schematic Diagram of Automatic Recognition Process Applied to Multiple-Channel Data

A fully developed analysis system based on the concepts outlined in Figure 1 could have several layers of techniques for pre-processing and processing. The great value of this approach is the multiple use of basic, extensive, and computationally demanding forms of information.

Interdependencies of different inputs and outputs of Figure 1 present a highly complex structure of information for implementing and evaluating. Many of the development problems will certainly be in the area of matching image properties to meteorological parameters, which if done in a very general way will be extremely complex. Judicial simplifications, especially in the early high volume states, are an important part in bringing this system into being.

Since the system must operate continuously and must also be able to incorporate occasional modifications in data and techniques, the best automated image analysis system appears to be one specifically designed to be adaptive. The automated aspects of this endeavor are being developed from first principles under the name artificial intelligence. Books by Carne,[13] Banerji,[14] Mendel and Fu,[15]

13. Carne, E. B. (1975) Artificial Intelligence Techniques, MacMillan and Co., London.

14. Banerji, R. B. (1969) Theory of Problem Solving, Elsevier Pub. Co., New York.

15. Mendel, J. M., and Fu, K. S. (Eds.) (1970) Adaptive, Learning and Pattern Recognition Systems, Academic Press, New York.

and Nilsson[16] give a broad view of this field. Sampson[17] presents an up-to-date introductory survey. The adaptive approach to automated systems presented by Tsypkin[18] has the advantage that it is both developed in a mathematical sense and practical in the operational sense.

Pickett and Blackman[19] express some doubts that the current state-of-the-art of artificial intelligence is capable of handling the problem of satellite imagery analysis. They suggest that developments in this technology be monitored and point out that improvements in information extraction techniques will increase the potential of artificial intelligence technology.

From the standpoint of techniques development, the presence of a high degree of flexibility and adaptability in a system is very desirable since imagery analysis is such a broad and unpredictable discipline undergoing rapid growth. On the other hand, decisions, no matter how good, place restrictions on the system. This becomes then an exercise in trying to foresee the type of equipment future techniques will demand and at the same time trying to satisfy current requirements. The choice of solutions is difficult, but it merits considerable thought and interaction.

3.2 Operational System at AFGWC

The image analysis system at the Air Force Global Weather Central (AFGWC) is used in analyzing both video and infrared imagery from DMSP satellites and is referred to as the "3-D Nephanalysis Model". The primary purpose of the 3-D NEPH is to develop a three dimensional cloud analysis over large parts of the globe on a near real-time basis. An early version of the system now in use is described by Coburn.[20] A more recent description of the AFGWC satellite data processing system is given by Canipe.[21] Basically, as far as the extraction of information from the imagery is concerned, the model uses brightness means and variances for 25 × 25 nmi areas. The infrared channel is used to make assessments of cloud top heights. While the model was developed on simple principles, its implementation has evolved into a fairly complex system.

In view of recent activity in the development of new extraction techniques and successes that have come in processing data in other fields, it seems likely that the fundamental components of the 3-D NEPH model can be upgraded. But ways to greatly improve the data extraction process without making large functional changes are not immediately obvious. The mean and variance statistics could perhaps be supplemented with other statistical measures.[19] Consideration of higher moments and a complement of textural measures for classifying satellite imagery would be the first to evaluate along these lines. The problem with making such an evaluation

(Because of the large numbers of references cited above, they will not be listed here. See Reference Page 67 for References 16 through 21.)

objectively, as with any other approach, centers around the fact that we do not have an adequate set of input imagery and a corresponding set of desired meteorological parameters.

Pickett and Blackman[19] surveyed the image analysis requirements at AFGWC and also examined state-of-the-art techniques thought to be of potential use there. They identified Fourier spectral analysis as the most promising technique to adopt as a first step in upgrading the system. The object initially would be to obtain spectral measures and perhaps others to supplement the mean and variance values. Work in this direction is being done at AFGL and AFGWC on a model to incorporate a Fast Fourier Transform analysis into the 3-D NEPH.

As a general rule, the more closely spectral measures can be related uniquely, or even partially, to certain cloud characteristics, the easier classification will be and the more straight-forward the analysis will become. On the other hand, if information is spread over wide spectral regions, its analysis may be far from direct. It is yet too early to assess the eventual solution.

4. IMAGE INFORMATION

The previous discussions have been introductory and hopefully informative. Their purpose has been to provide the context and background within which the following developments may be made.

The overall direction of our considerations was set not by the scientific area of interest—meteorology—but rather by the requirement that the analytical processing of satellite data be automated. As previously mentioned, in the early, high volume stages of analysis the problems of redundancy reduction and efficient compression of digital imagery data are very appropriate concerns. This may be referred to as "image data analysis", "image analysis" or "data analysis".

Requirements in the area of communications have stimulated in the past few decades considerable work in "bandwidth reduction" or data compression.[9,22] Although much of this research applies directly to problems of data analysis, the difference in the driving forces behind the two areas of research (that is, between communications and data analysis) produces some subtle distinguishing characteristics that are important to recognize.

The primary goal of the communications specialist is to transmit messages from one place to another. His challenge is to code (or compress) a message efficiently, transmit it, and at the other end of the line regenerate (decode) the original as well as possible under a set of given constraints. The introduction of noise is one of the more important constraints.

22. Janant, Nuggehally S., Ed., (1976) Waveform Quantization and Coding, IEEE Press, New York.

On the other hand, in data analysis the goal is to reduce data volume while retaining information required for decisions. The ideal would be to reduce it to the decisions.

Even though tools and methods are similar, this difference in goals causes an incongruity in results as well as in criteria for evaluating those results. This will not be elaborated upon, however, an effort will be made to explain differences in views and approaches as the occasion arises.

In this section some basic concepts of information theory will be presented to shed more light on the processing problem. Readers not familiar with this theory may find it useful to consult an introductory reference[23, 24] to better understand the following.

Our approach to digital data analysis is based on and derived from elementary principles of information theory as they are applied to fairly simple situations. Some illustrations will be given in this section. In performing operations on imagery, it appears important to do so with the digital significance of the data foremost in mind. Reasons for this view and for developing an approach to image analysis around it will be given.

To demonstrate what is meant, suppose we take an image having 64 levels of brightness (that is, 6 bits/picture element) and synthesize it on the basis of brightness levels and spacial arrangements of these levels. Here executions are made on 1 of 64 levels instead of the 1 of 6. This does not take advantage of a logical property of the concentrated form.

A folk-game that goes far back in history brings this point out in a straightforward way. It is known variously as "Bar Kochba"[25] or "Twenty Questions",[24, 26] and probably has several variations. Suppose you are asked to guess a number which, for our purposes, is from 0 to 63 inclusive. Questions are asked until you isolate the correct number. There are no restrictions on how the questions are formed, only that they must have a "yes" or "no" answer. The object is to find the unknown number by asking as few questions as possible.

A poor strategy would be to ask if the unknown number is such and such a number and continue by direct elimination. On the average this would take thirty-two and a half questions to obtain the required information.

The best strategy is one that does not give individual credance to each of the possibilities. Only six questions are needed to obtain enough information to learn

---

23. Abramson, Norman (1963) Information Theory and Coding, McGraw-Hill, New York.
24. Young, J.F. (1971) Information Theory, Wiley Interscience, New York.
25. Aczel, J., and Daroczy, Z. (1975) On Measures of Information and their Characteristics, Academic Press, New York.
26. Bendig, A.W. (1953) Twenty questions: on information analysis, J. Ex. Psy. 46(No. 5):345-348.

the correct number. The algorithm for arriving at this solution considers the 64 numbers in binary notation. This is a six-bit word. Questions are formulated that will produce knowledge of the contents, 0 or 1, of each position of the word. The answer to the sixth and final question specifies the number being sought. In this system, six questions are both necessary and sufficient to obtain the answer.

The difference between numbers of digits of data and the significance of those digits in a statistical sense is important in considering the compression of data in general and in seeking efficient techniques for reducing redundancy in particular. That this is the case follows from the modern quantitative concept of information introduced by Shannon.[27] This report was published in textbook form[28] along with an introductory paper by Weaver. The mathematics of the theory have been advanced considerably in the subsequent three decades.[25,29,30]

4.1 Information Theory and Image Analysis

Shannon's "mathematical theory of communication" (also called information theory) has become an important part of probability theory in less than thirty years. There has already been tremendous progress in both mathematical and applied areas and there appears to be no sign of an end to the fruits of this young and dynamic branch of probability theory. Here, we will review some of the fundamentals of that theory and show developments relevant to the "preprocessing" problem. While definite mathematical solutions to the image coding problem are not within sight, solid constructive solutions based on theoretical considerations can be obtained. Those elements of the theory that are of interest in this connection will be discussed here.

Shannon has defined information in probability terms as a "measure of uncertainty". This concept developed in mathematical terms is consistent with our intuitive understanding of information as an increase in knowledge about something.

"Information" is sometimes said not to be related to "meaning" and readers are cautioned not to associate the two. But this is not exactly the case. It is something akin to a flip of a coin in which one bit of information is provided. This information might mean that one person owes another five cents or five hundred dollars or something entirely different. In any event, it is still one bit of information. Information is a concept analogous to that of number in this respect.

---

27. Shannon, C.E. (1948) A mathematical theory of communication, Bell System Tech. Journal, 27:379-423 and 623-656.
28. Shannon, C.E., and Weaver, W. (1949) The Mathematical Theory of Communication, Univ. of Illinois Press, Urbana, Illinois.
29. Khinchin, A.I. (1957) Mathematical Foundations of Information Theory, Dover Publications, New York.
30. Feinstein, A. (1958) Foundations of Information Theory, McGraw-Hill, New York.

In the course of a game of "Twenty Questions", the answers, "Yes" and "No", progressively lead to the correct answer. In the course of events there was a progressive transfer of information; the player found out something that he did not know before. Information theory addressed the problem of measuring this change in "what is transferred" or the change in "what is known" as the result of some process. This is the central idea of the theory: that we can represent what is initially known as well as changes which occur when other information is received. This provides a measure for "amount of information."

An information source produces messages which may be labeled according to order of production; say S_1, S_2, S_3, S_4, S_5, ... S_n represents the first n of them. These are known to this point, S_n. The others, S_{n+1}, S_{n+2}, S_{n+3} and so on up to some total number $N = n+m$, are not yet known. Even though we have no way of predicting exactly what a given message will be, we can still estimate probabilities of occurrence for the next message S_{n+1}, on the basis of statistics of previously observed messages. For instance, in the set of n messages S_i occurs n_i times, then the probability of its occurrence is $p_i = n_i/n$. For each different message in the sequence S_1 to S_n we have a probability estimate of its occurrence. Let's call them p_1, 1_2, 1_3, ... p_r where r is the number of different messages and their sum equals 1. Since messages S_i and S_j together occur n_i and n_j times on the average in n cases, the probability that a message will be either S_i or S_j is $(n_i + n_j)/n = p_i + p_j$.

When the S_i's occur independent of each other, the process is said to be a <u>zero memory source</u>. When S_{n+1} can be guessed more accurately by knowledge of previous messages such as S_n, S_{n-1} ... the source has <u>memory</u>.

A message S_i is a very general concept. Even for a very specific numerical sequence, there is a freedom in the selection of what is to constitute a message. It might be a group of numbers or some variable grouping of numbers. It will be thought of here as a one- or two-dimensional array of positive numbers of the same fixed length.

The non-random message selection by a source, and time or space correlations (memory) make it possible to condense data to smaller volumes without loss of information. Data in this form is said to be <u>encoded</u> or <u>compressed</u>. The size of the system required to handle these data is called the <u>channel</u>. It can be described as the medium used to transmit the signal from transmitter to receiver.

A measure of information for the zero memory source which satisfies a number of desirable properties was described by Shannon and is written

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \, . \tag{1}$$

The entropy, H, is considered a measure of the average amount of information per source symbol of the given sequence. It was shown,[27] that $H \leq \log_2 n$ and that

the equality holds only when the p_i's are all equal to $\frac{1}{n}$. That is, the entropy is greatest when the source selects symbols at random from the given set of symbols.

The simplest example of a zero-memory source is that of a two symbol (say, 0 and 1) source with respective probabilities p and 1-p. This source has entropy,

$$H = -p \log_2 p - (1-p) \log_2 (1-p) \text{ bits} . \tag{2}$$

If p = 0 or 1, H = 0; that is, no information corresponding to certainty that p does not happen (p = 0) and certainty that it does (p = 1). This system has greatest entropy when p = 0.5, H = 1 bit. Tossing a coin is an example of a p = 0.5 system with no memory. The outcome of a toss produces one bit of information.

As it turns out, systems with entropies less than one are, in this case, not making full use of the one bit allotted to each message.

The fundamental theorem for a noiseless channel makes a limiting statement for message encoding. For a source with entropy H (bits per symbol) and a channel with a capacity C (bits per unit time-distance) it is possible to encode the output of the source in such a way as to transmit at the average rate $\frac{C}{H} - \epsilon$ symbols per second over the channel where ϵ is arbitrarily small. It is not possible to transmit at an average rate greater than $\frac{C}{H}$.

This not only shows how well the channel is being used, but also, how much improvement could be obtained by more efficient encoding. The actual rate of transfer of information divided by the capacity of the channel is used as a measure of the efficiency of the coding system. *This is a dimensionless number ranging from zero* to one.

The zero-memory source model can be improved on when correlations exist between successive symbols and sequences of symbols. Markov sources have been used for these situations. An m^{th}-order Markov source is one for which the occurrence of a symbol depends on a finite number, m, of preceding symbols and is completely independent of earlier symbols.

The probability of the i^{th} symbol, S_i, occurring after some particular sequence of m symbols will be denoted by $p(S_i/m)$.

This conditional probability (neglecting image boundaries when considering scans of an image) will have a definite value for each possible previous sequence of m symbols. Also, each of the other S symbols will have a similar conditional probability. The entropy of this system is obtained by summing for all symbols over all previous m sequences,

$$H(S) = \sum_m \sum_i p(m, S_i) \log_2 \frac{1}{p(S_i/m)} , \tag{3}$$

$$= \sum_{m+1} p(m, S_i) \log_2 \frac{1}{p(S_i/m)} . \tag{4}$$

Summing over all m and i for sequences of m followed by S_i, or (m, S_i), is equivalent to summing over all possible m+1 sequences.

Shannon showed that

$$\sum_{m+1} p(m, S_i) \log_2 \frac{1}{p(S_i/m)} \le \sum_i p(S_i) \log_2 \frac{1}{p(S_i)} , \tag{5}$$

with the equality occurring only when the probability of a symbol is completely independent of all previous ones, that is, when the source is reduced to the zero-memory source. This result simply means that correlations reduce information. And since we are at liberty to go back and redefine S_i to include a large block of data, these results apply as long as there are correlations.

As mentioned in a previous section, encoding techniques based on information concepts are greatly on the increase in digital TV coding for purposes of transmission. These techniques which explore both spacial and temporal correlations have apparently not come into use for image analysis purposes in the area of remote sensing.

Entropies of channels and entropies of sources, and statement of how they relate to each other under various conditions, place a measure on the desirability of encoding. The question of how to encode to realize these legitimate gains in compression has not been resolved. This is still in the art stage. Most real problems involve such a wide range of requirements that the present state of the theory does not apply. That is, it does not apply to answering the question of the best design of a compression scheme.

The compression scheme mentioned earlier by Marggraf[2] was developed to some degree with information concepts in mind. The procedure codes a satellite video image in small blocks. A pattern about the mean of a box is obtained and one from a coded set put in its place. Thus, each block is reduced to a mean intensity code and a pattern code. This approach has apparently not been developed further for satellite image analysis.

In addition to this coding scheme, Marggraf made some entropy calculations. To fully appreciate his work it is useful to follow through on the mathematics of conditional entropies for a nearest neighbor scheme.

To discuss correlation effects on entropy, consider a two dimensional array of discrete values. Consider two adjacent values x and y. Let i and j represent the intensity of the variables x and y respectively. The probability of the i^{th} intensity symbol is written p(i). For a given image, p(i) = p(j). The probability of the joint occurrence of i and j is p(i, j). The joint probability $p_i(j)$ is the probability of the j^{th} symbol occurring for y, given i for x.

By definition we have,

$$p(i,j) = p(i)p_i(j) ,$$

$$\sum p(i) = \sum_{ij} p(i,j) = \sum p_i(j) = 1 ,$$

$$\sum_j p(i,j) = p(i) ,$$

$$\sum_i p(i,j) = p(j) ,$$

$$\sum_i p_i(j) = p(j) .$$

The following entropy measures are defined:

for the event x;

$$H(x) = -\sum_i p(i) \log p(i) ,$$

for the event y;

$$H(y) = -\sum_i p(j) \log p(j) ,$$

for the event x and y;

$$H(x,y) = -\sum_{ij} p(i,j) \log p(i,j) .$$

It can be shown[27,28] that

$$H(x,y) \leq H(x) + H(y) , \tag{6}$$

and that the equality is true only if the variables x and y are independent. This equation shows that information obtained from (transmitted by) a pair of symbols cannot exceed the sum of the information obtained from (transmitted by) each symbol separately.

From the definition of conditional entropy, we can write,

$$\begin{aligned} H_x(y) &= -\sum_{ij} p(i,j) \log p_i(j) , \\ &= \sum_{ij} p(i,j) \log \frac{p(i,j)}{p(i)} , \\ &= \sum_{ij} p(i,j) \log p(i,j) \\ &\quad + \sum p(i,j) \log p(i) , \\ &= H(x,y) - h(x) , \end{aligned}$$

and since Eq. (6) can be written $H(y) \geq H(x,y)-H(x)$ it follows that

$$H(y) \geq H_x(y) , \tag{7}$$

which shows that if there are intersymbol dependencies between adjacent elements, the knowledge of x decreases the uncertainty of y; which means that there is less information contained in y than if there were no correlations. This analysis can be extended to include more variables; however, the representations become very cumbersome and calculations very demanding of computer time with only a small increase in the number of variables used.

Marggraf computed two-event entropies for TIROS images having 16 intensity levels (4 bits) and obtained overall values of 3.0 for $H(x)$ and 1.84 for $H(x,y)-H(x)$. These figures indicate a strong x and y correlation which can be seen by applying Eq. (6). Since $H(x) = H(y)$ in this case, we have $1.84 \leq 3.10$. This explains the incentive for coding of data grouped into small collections (block coding) and coding of data on the basis of errors in probabilities of occurrence (predictive coding).

By far the most work done in this area is related to TV transmissions. Both spacial and spacial-temporal coding schemes, have been developed. Unfortunately, incentives almost entirely result from problems of transmission rather than requirements for data analysis.

Calculations of image entropies which show data redundancies, are minimum under the assumed set of conditions. By improving our knowledge of image statistics this minimum is reduced. In other words, redundancy estimates obtained from probabilities of messages and message combinations are conservative. And they are especially so if there are redundancies in the image statistical structure not reflected in the entropy estimate. As a consequence, the development of image coding procedures, rather than relying on precise mathematical methods, relies heavily on intuitive ingenuity. The idea being to cover those weaknesses which are not accounted for by "exact methods".

Procedures involving a considerable amount of statistics have serious operational limitations. Practically any change in a system, unforeseen and unaccounted for, can make the operating set of "statistical constants" inappropriate. For this reason, adaptive techniques and those requiring little fixed statistics are preferred to those that involve a large amount of statistical information on data structure.

Coding techniques have two parts that must compliment each other, (1) a mathematical-statistical-intuitive rational that provides the "what is being done", and (2) a physical system of electrical machinery or whatever, that provides the "how".

An important paper by Blasbalg and Blerkom[31] classified the operations which compress the output of a message source into two groups, entropy-reducing

31. Blasbalg, H., and Van Blerkom, R. (1962) Message compression, IRE Trans. Space Electron Telemetry, 8:228-238.

transformations and information-preserving transformations. Entropy-reducing transformations are described as irreversible operations which result in reductions in fidelity that are acceptable to the user. The other group, information-preserving transformations, are those which map an input sequence into a corresponding output set that contains fewer binary digits. In this case, the process is reversible. That is, the input can be reconstructed from the output. Reductions in data amount result from redundancies and the amount of compression possible is directly related to the amount of redundancy present. The stronger the correlations within the data, the greater the redundancy. In addition to pointing out that source coding can be separated into these two parts, Blasbalg and Van Blerkom pointed out that source statistics are usually not known well enough to design fixed systems around them. This led them to adaptive coding systems.

A more up-to-date view of the mathematical aspects involved in entropy compression is presented by Gray and Davisson.[32] The short-comings of the Shannon approach to compression are discussed as well as problems and recent results suggesting that a mathematical theory of compression may be within sight.

Data compression schemes developed for image analysis are based largely on intuitive considerations. The foundation of a mathematical development of this area[32,33] from the point of view of information theory is under way. Important research papers making contributions to the rapidly developing field of data compression are compiled in a book edited by Davisson and Gray.[9] Of the forty-six papers reproduced, only about three are directed toward the analysis of images. While it might not be accurate to say forty-three totally ignore two-dimensional data, the tendency to operate in one dimension is very strong.

Although information theory does not rest on a dimensional framework, its developments and applications have been very much one-dimensional in nature.

The concept of information in its formal setting does not denote or restrict the dimensional setting for its application. The theory introduced by Shannon applies directly to one dimensional sequences. Higher order dimensions are treated by forcing them into a quasi-one-dimensional form. Even the terms used in this "mathematical theory of communication" have a strong one dimensional association, for instance, signal, message, channel, source alphabet, and sequences to name only a few.

The limited application to imagery of ideas of information theory has resulted in efforts to make statements of two dimensional data in one dimensional formats. There is apparently no unique and natural way this can be done. But nearly all applications of data require that they be in a sequence. Data sequences are

32. Gray, R. M., and Davisson, L. D. (1974) A mathematical theory of data compression, Proc. 1974 Inter. Conf. Commun. 1974, pp 40A-1-40A-5.

33. Berger, Toby (1971) Rate Distortion Theory, A Mathematical Basis for Data Compression, Prentice-Hall, New Jersey.

transmitted, operated upon, and decisions are made from them. And when a summarization is made of a two dimensional piece of information such as a photograph, it is usually developed in a one-dimensional sequence of descriptors.

It does not appear possible at this time to obtain a satisfactory sequential theory of image information. But how about a two dimensional theory? One not forcing a dimension change may be possible. One might ask, "What good are the results if we need them in one dimension?" The answer to this is that we may be able to perform simplifying operations and apply extraction methods in a pure setting before forcing statements of information into a sequence.

4.2 Image Information in Relation to Experience

The discussions in the previous section made little reference to concepts of imagery developed from many years of experience. These concepts form a major part of our knowledge about satellite imagery and represent the basis for most objective analysis schemes. This section will bring out some of the basic elements of experience that relate to the information analysis of imagery.

It might appear odd that this section on empirical results follows rather than precedes the previous one on theoretical developments of information. The reason for this is that the basic observational aspects of imagery involving visual perception are fairly well known, but not from an information point of view as presented here. This makes it reasonable that empirical considerations follow rather than precede the theoretical section.

Image analysis is more art than science at this time. It is dominated by results that are founded on human evaluations not backed by logic but with experience. This makes it difficult, if not impossible, to know what (or who) is right or wrong and risky to confide in any criterion of judgment other than final numerical results.

To show how this experience dominates the image analysis scene, imagine that there were, all of a sudden, no means of producing displays of any sort of satellite images. Say we have digital data, and means to synthesize them but not display them. Our almost complete reliance on vision would then be obvious. This is a simple statement of the automated image analysis problem and points up the weakness of objective analysis techniques currently available to process satellite imagery data.

Yet, without reference to a long backlog of experience, avenues for the development of reduction processes are extremely limited. One way or another analysis techniques must incorporate a large amount of deductions which are, in a sense, analogous to those obtained from experience.

The difference should be pointed out between what has been called "image information" and the information being sought in satellite imagery which will be called "meteorological information". Meteorological information is any statement derived from satellite imagery that reduces uncertainty of meteorological parameters. This

definition would have to be further refined to be useful in practice, but it is adequate to make the point about the relative volume of image information and meteorological information. Even if we insisted on being very thorough about meteorological statements, meteorological information would be only a small fraction of image information. There exist what may be termed meteorological redundancies. For instance, spacial statements of cloud cover, such as "complete area", "western half of area", "increasing from 20 percent in the north to 50 percent in the south", are compact forms of very lengthy statements. It is here that legitimate meteorological redundancies lie. Getting at them, however, has proved to be an enormous problem, the heart of which is data volume. For very small images and with very large computers, brute force methods can be applied to obtain solutions. This, however, is not normally the situation. Data volume and computational time are therefore important factors.

The satellite image analyst has developed ways to make assessments of images mainly from experience. While it is not possible and perhaps not even desirable to duplicate actions of the human analyst with computer techniques, there is little choice but to try to explain, and where possible use his findings.

The work that stands out in this area called "cloud interpretation" is that of Conover[34] published in 1962. This paper develops guides for determining cloud types and coverage from satellite visual images. The term "interpretation" was apparently selected to convey the fact that there is a certain amount of individual variability involved. Conover's guides to cloud analysis or interpretation are in the form of diagrams of branching processes which convert assessments of satellite image properties into cloud type and coverage. Properties considered were: (a) form - such as, round, curved, elliptical; (b) pattern - such as, banded, non-banded, randomly spaced; (c) texture - such as, smooth, fibrous, not fibrous, (d) brightness; (e) structure; and (f) dimensions of patterns and forms. Conover remarked that most of these image characteristics can be determined reasonably well except for brightness. The importance of accurate brightness measurements was stressed and factors were discussed that can thwart interpretation of clouds if not appropriately accounted for: (1) illumination or solar elevation, (2) position of clouds relative to sun and satellite, (3) reflectivity of clouds.

The layout of the cloud interpretation guides is of interest from the standpoint of the kind of decisions required of a human analyst. The first judgment required is "Are the clouds cumuliform or non-cumuliform?" Next, "Are they banded or non-banded?" At this point there are four branches in the decision process. The next set of questions differ for each branch and are more specific in nature. Then, depending on previous answers (or decisions that were made), there will be one of

34. Conover, J. H. (1962) Cloud Interpretation from Satellite Altitudes, CR Research Note 81, AFCRL, 77 pp; and Supplement 1, 1963, 19 pp.

eight possible branches to choose among. After this set of questions to answer (or decisions to apply), there are eight divisions.

Next is applied, for all cases, an element size-spacing judgment. And the last decision, leading to a final determination of the type of cloud being studied, is one of brightness in four categories; dark gray, gray, white, very white. The information arrived at, at the end of this process, consists primarily of cloud type and coverage.

The scheme undoubtedly produces consistency of results and simplifies the training of analysts. But more importantly, it increases objectivity in the analysis procedure.

The development of a general multi-channel guide for human analysis of satellite data is a complex task, and none, corresponding to Conover's for the visual channel, has been made. The greatest aid to interpretation since the early satellite days has come through an increase in the types of data available, that is, through more imagery channels, and in particular a far infrared channel. But a problem for the analyst with IR information is that it depends to a large extent on brightness. Texture, size, and orientation can be handled with considerable accuracy and confidence, but not brightness. Actually, the problem is more than that of brightness. It also includes uncertainties introduced by small scale spacial variations.

In any event, the analyst of data from a number of different imagery channels is apt to base his cloud interpretation largely on subjective procedures and his own past experiences. The results are frequently open to question and challenge by other experienced analysts. Automated analysis techniques are also weak at this time, but the untapped resources of the computer in this area may provide a significant improvement in the future.

4.3 Processing Images in View of Information Concepts

A question may be asked as to what is the information in a six bit (for example) image? The answer essentially is: that which reduces uncertainty about the state of the six bit image. More than anything else, this is a statement defining "information". In a given frequency distribution information is that which tells what the probability would be of selecting a certain brightness at random. It does not, in general, give any information on how values are arranged over the image surface. It is entirely possible that a wide variety of different images could have frequency distributions of brightness alike. This could seriously hamper any attempt to analyse images by this means. What the limits are to the amount of information that can be obtained in this direction have apparently not been worked out.

It is important, although it's not always easy to do, to keep separate what is known from mathematics and what is known from experience. This point will be emphasized further in subsequent discussions.

What are the important elementary properties of an image? Most persons from experience would list a half-dozen or more properties that have gained general status in one area or another of analysis-tone, texture, pattern, form, etc. Are these elementary properties that make up complexes of imagery? Are they basic, fundamental, indivisible properties? The answer is no, although it must be said that this does not impune the dignity of these concepts and will undoubtedly strengthen and clarify their use.

What basic form conveys information uniquely? Does image brightness carry some and elemental areas carry some? These questions at first appear simple, but comprehensive answers are elusive. Experience with spacial resolutions and brightness levels lead some to conclude that these properties determine the fidelity of an image and hence may be said to be the basic factors in the measure of information content of images. It is true that greater spacial resolution and a greater number of brightness levels provide a greater capacity per unit area for recording information. But they cannot be split in two separate parts and given independent value as to information, for information is a statement (or a quality measurement) about the structural arrangement of data.

If these arrangements are independent of the scale at which they are viewed, then information content per unit area would be inversely proportional to size of elemental areas considered. Increasing resolution by a factor of two would increase the information content by a factor of four.

Although they cannot be treated independently, element size and the number of brightness levels can be considered together in their influence on data and information. The central question is: How do these two factors combine or interact from an information standpoint? Notice that a given area with a given data rate leaves open, to a large extent, what these two values are (but not independently). For instance take an area, A, and suppose there are, R bits of information within it. What is: (1) the resolution of the elements in A, and (2) the number of bits for each of these (assuming they are all the same)? They are determined only within certain wide limits. On the low resolution side the limit is a resolution of A with R bits. On the high resolution side the limit is R elements within A having one bit each.

To further examine the question of data and information in an image or an elemental area of an image, consider the following situation. Suppose we have a high resolution image with both small elements and a large number of brightness levels, say 64. We could simplify these data by taking averages of adjacent cells and using these in place of the original values. If we average 3×3 elemental cells, the cell reduction will be by a factor of nine. But what happens to the number of levels of brightness? First let us see what the maximum brightness resolution would be if it were required that all the information be retained. The range of levels for this system is from 0 to 576 (9 times 64). This is the number of different average

brightnesses possible (actually 576 + 1). This is slightly more than 9 bits of information. If we did not bother with coding and used 10 bits to represent the data, the condensation of digits resulting from the decision to smooth in this fashion is 54:10 or 5.4:1. If it were agreed that 6 bits would be adequate the ratio would be 9:1.

The 9 (plus) bits in the example above would be relevant and meaningful only for exaggerated hypothetical conditions. Then, by what criteria can fewer bits be retained? (This problem arises in other forms, such as, how does one optimize the selection of element size and number of bits of brightness?) A major part of the problem requiring some sort of statement are: (1) image statistics, (2) required information. Answers to these questions are ordinarily arrived at from experience and "trial and error".

Back to the smoothing example. Suppose the image field varied very slowly as related to the 1×1 and even 3×3 element sizes. And suppose the 6 bit brightness values are highly significant but with some small random errors, and that accuracy is required. Under these conditions, the 9 (plus) bits are needed. This is just a hypothetical case to illustrate the effect of image statistics and required information.

But, in general, this is not the usual case. As a rule, when spacial smoothing is performed, cutting back on the significance of digits is justified. For most natural images, if they are smoothed spacially, it would be unreasonable to retain all of the resulting brightness resolution. The resolution of both should usually be of commensurate levels.

Additionally as a rule in spacial smoothing, the number of significant bits of brightness increases, but ordinarily no more than the number of bits (or brightness levels) of the original field are of any use. For instance, in the case above, rather than using 9 (plus) bits or 10 in the smoothed version, one is justified in cutting back to 6 for most natural images.

This rule can be expressed in other terms. If it is required that spacial resolution be relinquished through smoothing, the resulting increase in brightness resolution can also be relinquished. This empirical rule has not yet been verified mathematically. It falls into the same category as those which make summarizations based on spacial correlations.

It is assumed in the above that the original image has been optimized relative to element size and number of brightness levels for some given purpose. It is also assumed that this same purpose is the objective of the smoothed version. Mathematical study and verification in this direction are needed.

An initial requirement or step is that of developing mathematical statements of imagery for the purpose of developing reliable and general results. This is very important in its own right, for it is the beginning step in any mathematically rigorous treatment of imagery.

Before closing this discussion on image information, consider the following questions which elicit a different viewpoint. What constitutes information of an image? What part does brightness, texture, pattern, form, etc., have in terms of information contained in an image? These questions are meaningful only to the extent that the terms themselves are meaningful.

"Information" has been used here in the sense of Shannon in forming Information Theory. Of the many terms in image analysis denoting something about an image that could be labeled a type of information, (that is, the knowledge or measure of a characteristic which would remove some uncertainty) there are two fundamental ones which are brightness and texture. But this does not mean that they are distinct measures of information. By "brightness" is meant the sum of the radiant energy per unit area over some finite area. "Texture" refers to a measure of the variation of brightness over a finite area.

Since an array of brightness values defines an image, that is, specifies an image by definition, it follows that no other information forms can provide additional information. That is, any specification of an image by some property is either equivalent to a brightness specification or inferior to it. Consequently, terms used to describe imagery are a matter of convenience for pointing out certain characteristics of the array of brightness values. The key to this formulation is the requirement that an image be not only bounded in brightness but spacially finite.

What constitutes information of an image follows directly from, and is analogous to, the "one-dimensional" information. The information of an image element equals the amount of uncertainty removed by the brightness measure. This depends on the conditional aspects of the problem. Information Theory was developed and is ordinarily applied to a sequential passing of information from source to receiver, but here we have a problem that information does not form a naturally ordered sequence. In a sense we have as an image one huge source word. Such a word is a symbol of an instantaneous output of a source. These taken at equal time intervals form an information source. The totality of symbols of the source is its alphabet. Now, if the image is seen as one word, how is the information specified? The alphabet is enormous: $b \times 2^{nm}$ where n and m are the length and width of a rectangular image and b is the number of bits per "sub-word". How can the information be assessed or analyzed if the image is considered one word? The development of techniques to work within words may offer a solution to some of these problems.

5. A NEW IMAGE PROCESSING TECHNIQUE

Beginning with some simple considerations the development of a new automated means for encoding satellite imagery will be described. The main object of the

technique is to remove redundancies in the initial data while retaining spacial relationships within images.

A discussion of basic properties of finite image arrays is given in Section 5.1 in light of the foregoing assessments of information content. Then, in Section 5.2, various aspects involved in transforming an image to one bit per picture element are discussed. Two different views of the problem are given. In Section 5.3, a solution is presented that possesses many of the desired features such as being discrete, unique, and spacially symmetrical. The symmetrical feature, however, is subject to qualifications as will be noted. Finally, in Section 5.4, results of the use of this solution are given for high-resolution DMSP visual imagery. Comparisons of these images are made with one-bit images obtained by truncating the originals at intermediate gray levels. Calculations of an rms fidelity criterion are also shown which gave a quantitative measure of the "goodness" of the processing technique.

5.1 Finite Arrays

What constitutes information in an image? Image properties referred to by the photo-interpreter as brightness, texture, form, pattern, etc., convey information in a subjective, qualitative sense. So, it would appear that any analysis system of general form should be capable, at least in principle, of approximating such assessments. This is the primary incentive for image analysis, and is founded on the principle that since any real system has an information limit, it can be duplicated by a finite one. An interesting discussion of the finiteness of information is given by Van Soest.[35]

As stated earlier, an image will be viewed as a finite array of elements covering an area in the plane and with each of these elements will be associated a finite positive brightness value–in all respects this system is finite. A mathematician could justly argue that this places restrictions on what can be done (mathematically) within this system. He can conceive of a system that can not be accommodated by this finite one; however, all possible observations of such a system can be accommodated. But the important point from our standpoint is that there are advantages to this system not found in the continuous (infinite) one. This will become more obvious in what follows.

Consider an image in closer detail. Take one picture element and the brightness value associated with it. The conventional concept relating the two, associates a homogeneous gray tone, proportional to the brightness value, with the elementary area. But as for observations, the value is a summation of energy over the area.

35. Van Soest, J. L. (1956) Some consequences of the finiteness of information, Information Theory, edited by Colin Cherry, Butterworth Scientific Publications, London, pp 3-7.

But as for observations, the value is a summation of energy over the area. Without some other information than this, we are open to a much more liberal concept than that of a homogeneous gray tone. Any variation not conflicting with the system is as good as any other.

A few parameters may be specified for discussion purposes. Suppose brightness for a visible channel is coded into 64 discrete levels and is linear with a range from zero to that which includes the brightest cloud. The 64 levels can be expressed as a 6 bit word. If the darkest (lowest brightness) level observed lies above the lowest level in our system of representation, full value of the range is not obtained in the 6 bit form. This is correspondingly true also at the brightest level. To make it possible to account for this, the system can be re-scaled. Let "a" stand for the minimum level observed by the system and "b" stand for the maximum level observed. Both "a" and "b" fall within the 64-level brightness range interval.

Figure 2 illustrates how an image element can be viewed as a sub-array of bright elements denoted by ones, and dark elements denoted by zeros. One choice of arrangement of ones and zero is just as good as any other unless more can be assumed, inferred, or known. When nothing is known about the ultra-high frequency spacial variations of the system under observation, the assumption that the ones and zeros are distributed at random is attractive but not required. This demonstrates that an image is reducible to a 0-1 array without loss of information.

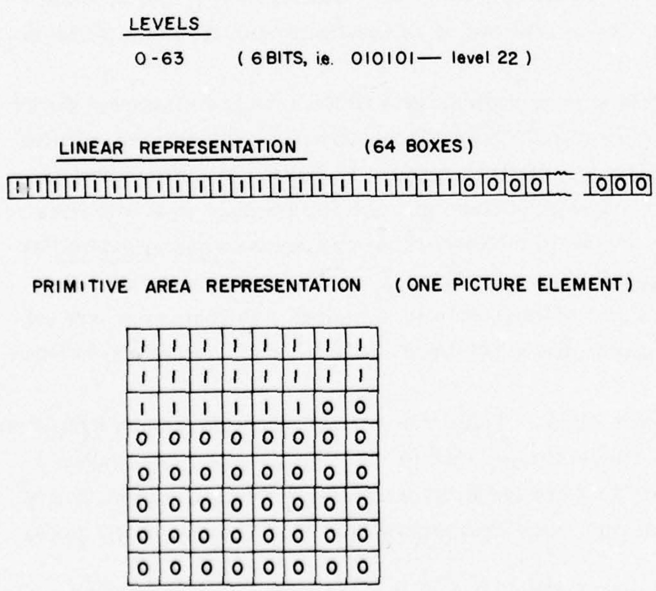


Figure 2. Primitive Forms of Numerical and Pictorial Representations

This system could be very large and unwieldy when it comes to implementing. There's a simple way around this difficulty, although it does retract some from the position already stated. Rather than obtain actual distributions from this abstracted view, consider for calculation purposes that each of these subelements is located at the center of the element proper. It will be shown how this concept is useful for manipulating imagery.

Note that by switching representations, and by taking care not to read anything into (any aspects not present in the original) those which are formulated and by assuring ourselves that both are equivalent in our interpretation, they both have the same information regardless of any differences in volumes of numbers. Also it is much more amenable to the problem of reducing redundancy than the traditional representation, and that is the main purpose of this abstraction.

In Figure 2, the subelements of the picture element can take on one of two brightnesses. "One" denotes high brightness, "zero" low brightness. The incrementing interval is the positive difference of these values. The system could actually handle one more level than the six-bit representation can, but that incongruence will not be explored here. It is sufficient to know that a finite image representation can be obtained which has only two levels of brightness.

The only additional limitation put on the system from that described initially is that brightness values are bounded on subelement scales both from above and below. In terms of a visual channel this makes it possible to take out wasted numbers used in the system below the brightness of the surface, especially the limiting low value of reflection from the ocean surface.

This finite array represents a two dimensional information source. Although the representation requires 63 bits, the entropy or actual number needed to carry the same information is only two bits (give or take a bit).

The objective in the following sections will be to obtain a scheme to encode these 63 bits into a one-bit representation. It will not become necessary to make an actual representation of these subelements either physically or geometrically. So, for all practical purposes, it is merely a mental construct to aid in developing a method. Nevertheless, even though a construction will not be needed, the representation is constructable (as opposed to an abstraction that has no physical representation).

5.2 The Problem of Transforming an Image to One Bit per Picture Element

As stated in the introduction, the primary problem considered in this report developed in connection with a requirement to compress special purpose satellite imagery into one bit per picture element for purposes of handling, computation, and storage. The solution arrived at is, however, somewhat general and makes the process of interest from the extraction point of view as well as that of redundancy reduction.

As seen earlier, the information capacity of the binary digits of satellite imagery is much less than what information theory indicates is the maximum capacity of the digits. This result provides the challenge to find a code that will permit simple, contacted statements of satellite images while retaining information. In particular, the search is for a coding procedure that will transform a standard six-bit image into one having a one-bit format. There is no guarantee that the entropy is actually this small (one-bit) in which case some information would necessarily be lost even if the coding procedure were completely efficient in storing information into one-bit. But, all indications are that the entropy does not exceed one by very much, if at all; therefore the challenge to pack one bit as full as possible remains.

Rather than be concerned about entropy at this point, consider the question of how well a one-bit image can convey information. The first reaction might be that such a contraction would be very limited at best. But this is not the case as can be seen by making some simple calculations of (1) the number of patterns possible, and (2) the number of brightness levels that can be realized over small image areas. For example, a three $\times$ three array of one-bit elements, collectively, can take on two^{9} configurations and represent ten brightness levels. These two realizations are, of course, not independent of each other.

In this section, the problem of reduction to one bit will be looked at from various angles; there are a number of interesting facets of the problem which can be stated in very basic terms. It was through repeated evaluations of these elemental parts that a route was found constituting a solution. Before the solution is outlined in Section 5.3, however, some mathematical aspects of the problem will be described in the following.

Let $G(x, y)$ represent the brightness values of the original image array and $\psi(x, y)$ a one-bit approximation to it. The one-bit array will frequently be labeled with symbols 0 and 1. The convention will be that 0 represents a low tone and 1 a high tone. They can be thought of as black and white tones respectively. The exact specification of these two tones represents a negligibly small number of bits in comparison to the bits of the images being considered. Understanding of this simple aspect is fundamental in moving toward a workable code. The 0, 1 labels carry the image information. In a previous section it was pointed out that information derives from arrangement and that without variability in a structure there is no information. These labels permit variability in its simplest form.

Information in the discrete one-bit array does not depend on physical dimensions but on the spacial arrangement of the two levels of brightness represented by zeros and ones. In coding to one-bit, therefore, the interest is in preserving the essence of these information structures while at the same time arriving at an abbreviated statement of the image. Since brightness represents the basic building block, the conservation of areal brightness will preserve information.

The problem to consider can be stated in general terms as: find a process to obtain a one-bit image from a conventional image in a way which minimizes the square of the difference between the average brightness of corresponding arbitrary areal collections of picture elements.

This is a least squares fit of $G(x, y)$ with the bi-variate $\psi(x, y)$ under the conditions stated. The rule concerning average brightness discussed earlier applies here in that the areal averages of different sizes are not weighted. All differences are minimized as they appear. The term "areal" in the statement of the problem can be dropped as far as calculations go. It is there to assist in the visualization of the problem. A direct implementation would have to put a limit on the number of these collections and various areal collections could serve this role.

It is obvious from this statement of the problem that a very large number of equations with $n \times m$ (number of points in image array) unknowns, specifically $\psi(x, y)$, can be written. And such an approach is clearly out of reach computationally for all but the smallest of images unless some very effective simplifications are made.

Some appreciation for the magnitude of the numerical problem involved in obtaining a solution to agree with these statements can be obtained by considering some corresponding collections of image elements. For each point, a comparison can be made to give

$$[G_1(x, y) - \psi_1(x, y)]^2 = Z_1, \tag{8}$$

where the subscript 1 refers to 1 by 1 elemental boxes and Z_1 is a part of the quantity to be minimized. For larger areas, this can be followed by

$$[\overline{G_2(x, y)} - \overline{\psi_2(x, y)}]^2 = Z_2 , \tag{9}$$

$$[\overline{G_3(x, y)} - \overline{\psi_3(x, y)}]^2 = Z_3 , \tag{10}$$

$$\vdots$$

$$[\overline{G_k(x, y)} - \overline{\psi_k(x, y)}]^2 = Z_k , \tag{11}$$

where k is some large number that Z_k becomes so small that going further would not improve the solution.

The main area of consideration lies at intermediate levels. For when $G(x, y)$ is either very small or very large there is no question what the appropriate $\psi(x, y)$ should be. For $k = 1$, little can be gained, since one selection of variables is about as good as any other.

The object in the above is to select the bi-variate variables to minimize the sum of the Z's, that is,

$$Z_{min} = \sum_{k=1}^{n} Z_n . \tag{12}$$

For k = 2 and intermediate values of G(x, y), $\psi(x, y)$ can be selected to give small Z_2; namely, half low tone and half high tone averages to produce an intermediate tone.

As k increases the ability to represent more levels increases but the ability to represent spacial change decreases and consequently the bivalent variables are specified by intermediate values of k. The actual selection of the $\overline{G(x, y)}$ and $\overline{\psi(x, y)}$ collections is open to choice or (if the problem statement is taken as it stands) to the definition applied to "arbitrary collections".

In any event, the primary objective has the following features: (a) one-to-one transformation in terms of picture elements; (b) conservation of brightness over large areas and approximate conservation over small areas; (c) assignment of one of two tones to each picture element in such a way as to minimize, according to the method of least squares, differences between corresponding arbitrary collections. Arbitrary in this case means selections of different size without a bias toward any one of more sizes.

This approach to the definition of the basic mathematical aspects of the problem can be expressed mathematically as,

$$\sum_x \sum_y \left[\overline{G_n(x, y)} - \overline{\psi_n(x, y)}\right]^2 = Z_{min} . \tag{13}$$

This expression leaves the question of weighting unspecified.

While this is a relatively clear statement of the problem, it does not lead to a means of solving it. It does provide, however, additional insight, gives an indication of the complexity to be dealt with, and also indicates a way to measure the "goodness" of a solution.

In view of the results on finite arrays in the previous section, a physical statement of the problem can easily be developed. This will indicate how the problem can be approached in incremental steps.

Consider an array and a sub-array as discussed in the previous section. Let each sub-element either contain or not contain an object (analogous to 1 and 0). Each array has associated with it, say at its center, a place for a container (analogous to a grid point). These containers hold only n objects and there are just enough of them to hold all of the objects in the array.

The problem is to identify the locations at which to place the containers which will minimize the work required (total transport distance) to fill them each with n objects.

This is a simple straight-forward problem, and obviously must have a solution. But, as with many problems in discrete optimization, beneath this plain and simple looking structure lies enormous complexity.

That these two statements or fomulations describe essentially the same problem is certainly not obvious on the surface; but, that they are related to a high degree can be ascertained by closer inspection.

Consider both problem statements in relation to the representation (in terms of subelements) described in the previous section. Both require the formation of subsets under conditions which require a minimum difference between the original set and the solution on an areal basis.

Approaches to both of the problems seem to require greatly simplifying assumptions or decisions to concentrate on some aspects while neglecting others. The one dimensional version of the problem provides considerable insight here. It is seen that by requiring an exact solution, much effort can be spent after an adequate solution is reached in forcing an exact solution. This suggests that a penalty system be added so that the distance from a solution can be compromised for operational purposes. There are two separation distances of importance here: the distance from the original data and the distance from the ideal solution.

In the final analysis, the object is to find an efficient process which can be applied to image data that will give a close approximation in the sense noted above. This is the topic of the next section.

The statements of the problem indicate requirements that can be imposed which are advantageous in that they are restrictive in the sense of obtaining a unique solution and are not confining in terms of method. Before proceeding to the next section these major characteristics can be enumerated as follows:

(a) Discrete and finite system in all respects,

(b) One-to-one transformation in terms of picture elements,

(c) System bounded both above and below,

(d) Conservation of image brightness over large areas,

$$\sum_x \sum_y G(x,y) - \sum_x \sum_y \psi(x,y) = 0,$$

(e) Minimum difference between original and solution for arbitrary areal collections,

$$\sum \sum [\overline{G_n(x,y)} - \overline{\psi_n(x,y)}]^2 = Z_{min},$$

(f) Symmetry of any actions within the image plane to prevent directional biases,

(g) Image brightness information and scale are interrelated such that one has meaning only in relation to the other; and this is different from that expected from combining elements.

These basic characteristics of the problem appear to be both consistent and complementary.

5.3 An Algorithmic Solution to the Problem

What should a good solution to the transformation problem look like over an image surface? In considering large dark ocean areas in a satellite with visual image, for example, the solution should be uniformly the lower of two brightness values available. Over large bright cloud areas it should be uniformly the higher of the two. For an expanse of cloud with brightness midway between these extremes, there should be about half-and-half of the two levels distributed at a high spacial frequency. The "high spacial frequency" requirement insures agreement of averages over small areas for the situation being represented. For areas of irregular brightness, the solution should contain mixtures of the two brightness levels reflecting the irregularities present. If averages were taken to make a coarser resolution image the solution, if it were a good one, would approximate the corresponding averaged image obtained from the original. And this approximation could be expected to improve with increased smoothing (averaging resulting in reduced spacial resolution).

A truncated version of the original image does not give results of this nature. Intermediate values are falsified either one way or the other and large areas can be totally misrepresented.

The significance of this problem in a numerical sense demands an incremental approach to its solution. What is required is a process that transforms $G(x, y)$ to $\psi(x, y)$ which makes small changes to the former while making progress toward the latter.

Although other solutions may exist the one described below was arrived at by a process of heuristic reasoning similar to that described by the great mathematician-teacher, George Polya.[36,37,38] A more recent reference to this topic, which reflects the great activity in optimization of the last few decades is an essay by Koopman.[39]

The eventual optimal solution (if and when it is obtained) may be quite complicated for actual use, but it would nevertheless be very desirable for evaluating

(Because of the number of references cited above, they will not be listed here. See Reference Page 67 for References 36 through 39.)

computational approaches such as the one presented here. The problem can be stated mathematically in several ways and one or more may yield to analytic solution. One such formulation, which differs basically from the completely discrete one discussed here, has some desirable features. The brightness range could be considered continuous making the image a piece-wise continuous function of x and y. This would permit the use of classical techniques of analysis. But in the end, results would have to be mated with finite observations. In fact, any solution should be closely associated with the physical form of digital imagery to be of practical use. This feature stands foremost in our solution and it is difficult to imagine a treatment of the problem having a stronger association between method and data.

There is another prominent and desirable feature contained in the configuration of the problem. The completely discrete-finite statement confirms the existence of a solution. It is easily seen that not only systems $G(x, y)$ and $\psi(x, y)$ are finite but that there are a finite number of ways that $\psi(x, y)$ can be configured to approximate $G(x, y)$. There must also be one arrangement for which there are none better (for some given criteria for better). The preceding statement is worded in this way because of the possibility of more than one such optimal arrangement.

The word "finite" has been used here in a mathematical sense. A few calculations of the number of arrangements possible for, say, a small one-bit image, will reveal that the definition does not place a limit on the size of finite numbers.

Further assurance of flexibility of a one-bit image system for representing information can be obtained by noting the number of brightness levels possible over $n \times n$ elements for various n. For $n = 1$ there are 2 levels possible, for $n = 2$ there are 5, and for n in general there are $n^2 + 1$ brightness levels possible. (This is consistent with the earlier result that a six-bit image (64 levels) can be represented on an 8×8 sub-array with one subelement not needed.)

The foregoing definitions and developments have set the tone for the following approach. It is first of all discrete in that units of brightness are not broken into parts. It was felt that this approach would be most direct in terms of basic computational structure. For the conventional computer this may not be the case. The process begins with the original image, $G(x, y)$, and operates on it with the following restrictions: (1) conserve local brightness averages; (2) get the most change toward a solution with least amount of change in local brightness; (3) retain as much symmetry as possible in the operation. If this last restriction was not included then there would be no assurance that the process would not favor some direction. To avoid extensive bookkeeping in the computations to insure symmetry the requirement is imposed at the basic level. A random selection process might be used here, however, a definite and direct operation was preferred.

Consider an element and its eight nearest neighbors. The transformation from $G(x, y)$ to $\psi(x, y)$ can be considered as a step-by-step process. Small steps are made to permit a gradual adjustment. Large steps would force a solution dependent on how the process is applied. The overall brightness does not change since a procedure is adopted in which an element of brightness at one location is subtracted while another element is added at another location.

The requirements for a solution can be seen more clearly if the scheme shown earlier of an image representation consisting of two tones is used. A picture element in this case is viewed as an 8×8 array of subelements. The object then is to transform the original in this form to an arrangement that has a small number of bright elements and a large number of dark elements (representing smallest energy level), or a large number of bright elements and very few dark elements (representing largest energy level). Then there are essentially two levels of brightness (for example, one-bit representation). To obtain this, bright and dark elements may be interchanged according to some rule that satisfies the requirements discussed earlier. The interchanging of elements insures an overall conservation of brightness. The distance over which the interchange is made is a measure of the error (not actually error but deviation) resulting in the change. The object is to keep these as small as possible while at the same time causing the intermediate brightness values to either decrease or increase to the limits of "a" and "b".

Consider the requirement that the overall brightness of $\psi(x, y)$ should equal that of $G(x, y)$. Rather than normalize $G(x, y)$ or obtain a probability field, we can work with the given field and make transfers of brightness values, adding an amount at one location while subtracting it from another. The object is to find a process that will satisfy this and other requirements.

In statistical sampling random errors can be reduced without limit by taking a larger and larger number of observations. This holds whether the distribution is normal or not as long as it has a finite mean and standard deviation.

Averaging of a number of elements in an area can be viewed as a sampling process to reduce random errors of brightness. The variance is given by $\sigma m = \frac{\sigma}{\sqrt{n}}$. This result from statistics indicates that random errors can be reduced without limit by using the mean of a larger and larger number of observations, provided that the distribution of errors (whether normal or not) has a finite mean and standard deviation.

To create a situation where areal brightness values remain essentially unchanged, units could be transferred between adjacent boxes. That is, units could be taken from smaller values and added to larger values. This would increase variance and at the same time tend to conserve brightness values over small areas unless there were long runs of continuously increasing values. In any event, this

would bring the image spectra closer to the goal of being "two-valued" while approximating the multiple-valued one given. But can transfer up a gradient for long distances be avoided? Consider what can be done in a one-dimensional setting for ease of comprehension. Four values are needed as follows:

$$x_1 \quad x_2 \quad x_3 \quad x_4$$

and the problem is what to do with x_2 and x_3 in view of all four values.

Suppose $x_2 > x_3$ (the same results will apply if $x_3 > x_2$ but in reverse); the previous process would make an increase in x_2 at the expense of x_3. Now suppose x_1 is less than x_2. There is no objection to the action taken. But not if x_1 is greater than x_2. To prevent undue transfer of brightness values, consider the combined relationship of x_1 and x_3 to that of x_2. Now transfer <u>from</u> x_2 to both x_1 and x_3 if $[(x_1 + x_3)/2] > x_2$, and transfer <u>to</u> x_2 if $[x_1 + x_3)/2] < x_2$. In other words, if the center value is low relative to the average of its two neighbors, then it is decreased by two units and they each are increased by one. If the center value is high relative to the average of its two neighbors, then it is increased by two units, one from each neighbor. This procedure discourages shifts of brightness values and promotes a local persistence in brightness.

This decision criterion may be recognized as the basic finite difference form of the Laplacian operator. The development has not taken grid distance into account. The object is to keep the procedure integral.

Extending these ideas to two dimensions we arrive at the algorithm shown in Figure 4. The iterative procedure shown there causes image brightness values to separate in a way which tends to preserve averages over small areas. In fact, for a given application of the process to a point there is no net change in brightness for the 3 × 3 box under consideration. Picture elements are processed sequentially, left to right, top to bottom. The brightness of an element G_o, is compared with the average of the four nearest neighbors of the "cross" in Figure 3. If it is larger, it is increased at the expense of the four neighbors. If it is smaller, it is decreased to the benefit of the neighbors. No value is decreased below a minimum level, "a", or increased above a maximum level, "b". When finished with the "cross" processing, the same point is processed in the same way for the "diagonal" setup of Figure 3. The incremental change is "c". The asterisk means exit to "diagonal" if in "cross", or advance to the next point if in "diagonal".

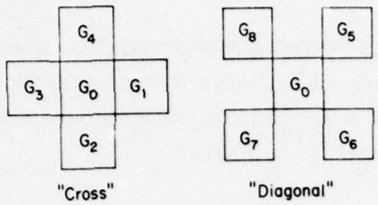


Figure 3. Representations of Gridpoint Designations

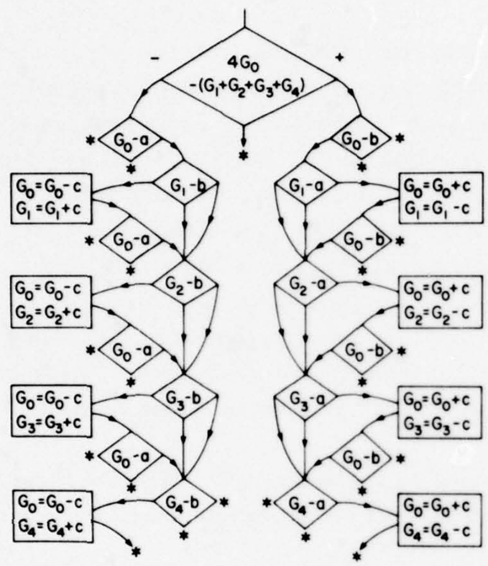


Figure 4. Algorithm for Separating an Image Into Two Levels. The G's refer to gridpoint designations (Figure 3); the asterisk means exit (see text for details); "a" is a lower bound, "b" an upper bound, and "c" is an increment of brightness

An interesting point may be made here in terms of changing or losing digital information. If a process is reversible and can be brought back to the original with the same interpretation as at the beginning, then the process preserves information. Consider the algorithm in Figure 4. When it is separating an image and is not in a state where grid points have reached either a maximum or a minimum value and when $4G_0$ is always unequal to the sum of the four neighboring points, the process is reversible. This includes the whole iterative process: "cross" process, "diagonal" process, point-to-point, line-to-line, and repeated iteration. This gives some assurance that the separating process of the algorithm tends to conserve information. But this does not mean that the transformed version has the same information in the transformed sense.

It has been found that four passes are enough to separate brightness values into essentially two distinct levels. At this point of separation, the array is truncated to a one-bit field. As a substitute for an optimal analytic solution with which to compare results, the measure described in the previous section will be used to evaluate the differences between areal means of $G(x, y)$ and $\psi(x, y)$.

The constant "a" can actually be taken as a variable if there is a data bank available on background brightness. In this case, the image can still be represented by one bit. In a visual presentation of the zero-one single-bit data, the ones could take on the brightness constant, "b", and the zeros take on the background brightness, from the data bank. If there are many images over the same background brightness, this makes the process even more attractive for handling the data.

The foregoing indicates that it is very desirable to separate details possessing a small amount of information from those with much information and then to treat them accordingly in the analysis. This is a feature that should prove very useful for certain applications. For instance, land patterns in the visible data have a large amount of detail but a small amount of information. Recall that information refers to the removal of uncertainty. The background brightness details remove very little uncertainty when they are already known from previous observations. On the other hand, they are needed for delineating cloud information. This situation illustrates some of the differences in the concepts of data and information.

5.4 Results of Algorithm Applied to DMSP Visual Imagery

The algorithm described in the previous section has been tested on six cases of very high resolution (1/3 nmi) DMSP visual imagery. These images were in a six-bit (per picture element) format.

The object of the tests was to determine how well one-bit images produced by the algorithm capture the essence of visual cloud images. It was decided that this information would be obtained by comparing the one-bit images obtained with the algorithm, with one-bit images obtained by truncating the original image. This gives a visual comparison to supplement direct comparison with the original image itself. In addition, an objective means of evaluation described earlier has been applied and some calculations of rms differences presented.

In all cases the original data have been operated on and transformations are displayed in a one-to-one form, except for some enlargements of small areas to be shown later. Images were displayed and photographed on the AFGL McIDAS, which is a man-computer interactive data analysis system. Most calculations were performed on the AFGL CDC 6600.

Figure 5 provides an orientation to the scale of the 1/3 nmi data. The "whole mesh box" grid used at the Air Force Global Weather Central is shown superimposed on the McIDAS TV screen at the same scale as the 1/3 nmi data, that is 80×80 picture elements forming one 1/8th mesh box. The boxes are 25 nmi on a side. The McIDAS screen outline is shown by a heavy solid line. This screen can present five of the eight rows of 1/8th mesh boxes. It could actually handle

six except that some scan lines at the bottom are used for identifying labels on the pictures.

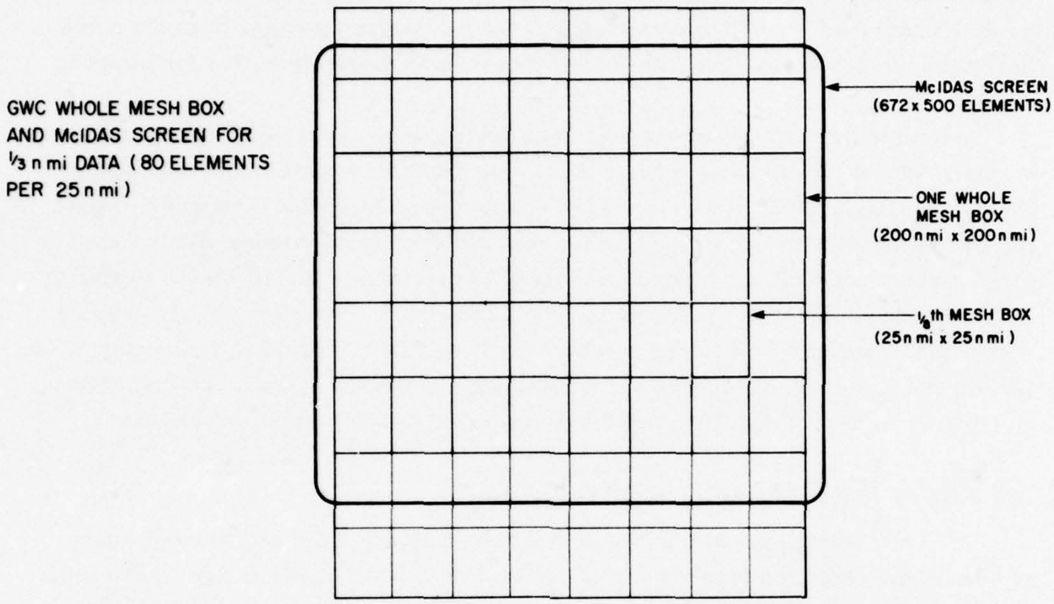


Figure 5. Comparison of Standard GWC Grid and McIDAS Screen for Very High Resolution Imagery

The DMSP visual imagery provides a good test for the algorithm since it has a wide variety of image characteristics through the range of darks to lights. It has regions of high contrast and low contrast with considerable variation between. There are also various sizes of "objects" against dark or moderately gray backgrounds.

The original image presentations of the six cases used for the tests are labeled A through F in Figure 6. These samples, all from low latitudes, show a fairly wide range of cloud types and variabilities of brightness. Most of the pictures are of ocean areas except Case C and part of Case E. Brief descriptions of the six cases follow:

Case A shows an area of low fair weather cumulus that exhibits a wide range of cloud cover and cloud element size.

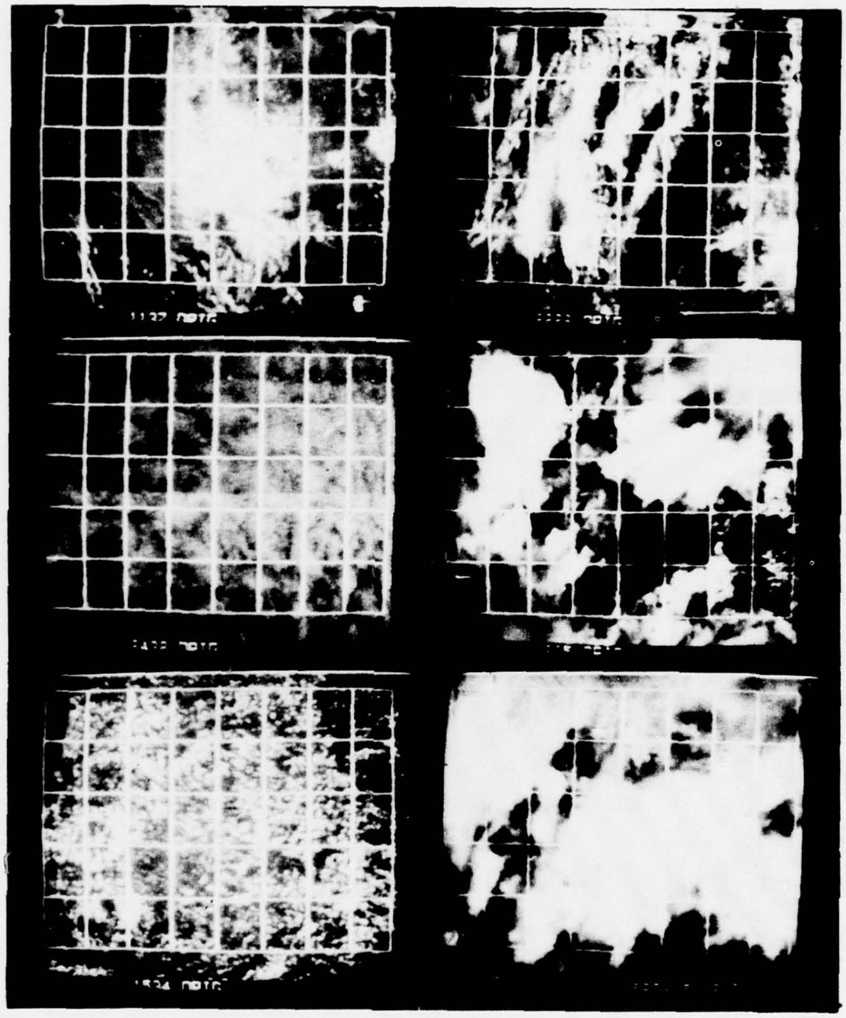
Case B is of a cellular cloud pattern. This type of cumulus cloud pattern frequently occurs over large areas of oceans in middle and low latitudes. Such patterns are not nearly as bright as those that result for solid cloud, not because of

low reflection from the cumulus clouds, but because the clouds are smaller than the footprint of the sensor. They do not fill the field of view and the measurement is an average for cloud and ocean. An instrumental level change occurred about halfway through this image.

A

B

C



D

E

F

Figure 6. Originals of Very High Resolution Visual Images (with 25 nmi grid overlay) Used for Calculations. The six cases are indexed A through F

Case C shows cumulus over land (Honduras and Nicaragua) with a wide range of sizes. These clouds are partially a response to solar heating of the earth's surface and lower levels of the atmosphere.

Case D shows a variety of clouds; cumulus, cirrus, and areas of thunderstorm convection. There is a large range both in cloud element size and brightness.

Case E has even larger areas of convection. The one in the upper left of the picture is about 50 nmi wide (E-W) and 75 nmi long (N-S). The other cloud area to the east is about the same size. It appears to be made up of a number of thunderstorm areas. Cirrus blow off from an area almost off the picture is shown in the upper right corner. There is a nearly circular cloud free area with a 30 nmi diameter in the lower central part, other cloud-free spaces are seen in other areas.

Case F shows a storm area with overlying cirrus. Dark gray areas are cirrus without bright clouds beneath. Black areas are of the ocean surface without any significant amount of clouds. There are variations of brightness over the storm areas that suggest the storm is made up of smaller organized clouds.

These cases represent a varied assortment of clouds of the type that an automated processor will have to contend with routinely. Yet, there are many other types that could be added, such as a variety of stratiform clouds, more layer combinations, and patterned clouds such as bands, cells, streets, etc. A comprehensive data set with ground truth is very much needed for performing classification experiments and for comparing schemes of all sorts. Such a set would be useful for generating structural statistics as well. This, however, is a bigger task than it initially appears for there are important and difficult questions that must be dealt with. Advancements in automatic processing of satellite imagery will probably be slow until this important job is completed. For our purposes, the set of six images described above is sufficient for making an appraisal of the technique for converting standard imagery to one-bit-per-picture-element imagery.

Since the McIDAS system uses eight bit words as a standard, a conversion of the six bit imagery data to eight bits was performed by multiplying it by four. All executions however, were made in integral units as if they were six bit words, that is, in units of four.

The decision to keep calculations on an integral basis was made for simplicity and especially for application to computers that operate on word bits in a basic way. Also, this form makes for great simplification of hard-wiring logic over the conventional numerical approach. In short, there are apparently no significant sacrifices of information in maintaining integral calculations, but there are some definite gains.

From the standpoint of the standard computer, this integral word-logic approach makes demands <u>not</u> designed into them, and does not use features that <u>are</u> designed into them, in a very efficient manner. For instance, standard computers are built

for multiplication, division, and so on, in terms of big words; some computers have words as large as 60 bits. This type of capability is not used in the solution described here. Rather, words are restricted to six bits, and the subtraction and addition of one unit from or to them. Most of the calculations involve shorter word lengths than six bits, and a large fraction involve words of only one bit.

Initially, experiments were made with the five point (G_0, G_1, G_2, G_3, and G_4) "cross" setup alone, using eight and ten iterations. This appeared to be about the right number of iterations to get good separation, however, the images did not measure up in other respects. These trial runs with only the "cross" version produced irregularities in the final stages of iteration which apparently resulted from the non-symmetry introduced by using the "cross" process alone. The "diagonal" part was added and the irregularities no longer were observed. This part of the calculations was made to follow immediately after the "cross" version and before advancing to the next point (G_1, G_2, G_3, G_4 becomes G_5, G_6, G_7, G_8). There was no accommodation made for the difference in distances of these two sets of points from G_0. Distance was not a part of the development; the requirement to keep the calculations on an integral basis was judged more important. Using both the "cross" and "diagonal" processes together increased convergence by about two times so it was possible to drop back to four iterations.

One possibility of squaring off transport non-symmetry due to distance differences is by applying the cross and diagonal parts at different frequencies. Another aspect in this same vein (that is, falls in the category of symmetry) is that of random application of the algorithm rather than sequential application. Experiments have not yet been conducted in either of these areas.*

An experiment was run with the value of a = 0, and it was found that this caused too much shifting (and consequently computer time) over the ocean areas with little or no increase in product accuracy. It was then decided that the ocean brightness or surface brightness would be a better value for "a" in general.

This algorithm has been applied to the six samples of DMSP visual data shown in Figure 6. Results of the calculations in image form are shown in Figures 7A, B, C, D, E, and F. In each of these, the original is given in the upper left and below it are three quantized images; the truncation levels for these from top to bottom are 25, 30, and 35. The image in the upper right is the result of four passes of the algorithm as described above. The three images below it are truncations of that image at the same levels as the picture to their left (25, 30, and 35).

---

*As for the non-symmetry of the "cross" and "diagonal" processes, it should be observed that a hexagonal array of data would clear up this incongruency, however, data are not usually in that form.

The one-bit images resulting from iterations of the algorithm will be referred to as "bisected" images from here on while those obtained by quantizing the original images will be referred to as "truncated" images. The image obtained from four iterations of the algorithm will be called "fourth iteration" images.

By this nomenclature, the layout of the sample images which appear in Figures 7A–F is as shown in the following diagram.

Original	Fourth Iteration
Truncated (level 25)	Bisected (level 25)
Truncated (level 30)	Bisected (level 30)
Truncated (level 35)	Bisected (level 35)

The one-bit images were presented on the McIDAS screen at the levels of 40 (10 in six-bit notation) for the dark and 200 (50 in six-bit notation) for the light. (The size of the full screen is 672 elements wide and 500 elements nigh. Twenty-five lines were used for identifiers giving a viewing area of 672 × 475 elements or about 224 × 157 nmi for the 1/3 nmi data.) The bisected images when viewed on the McIDAS TV screen have a flutter which is apparently caused by time differences in registration. Photographs of the screen were taken over an interval of 1 sec in order to average out these quick changes in brightness.

The truncated images are shown as controls and different levels are given since this is an important variable (a constant in a working system). The truncated images show the results, or consequences, of this straightforward reduction to one bit.

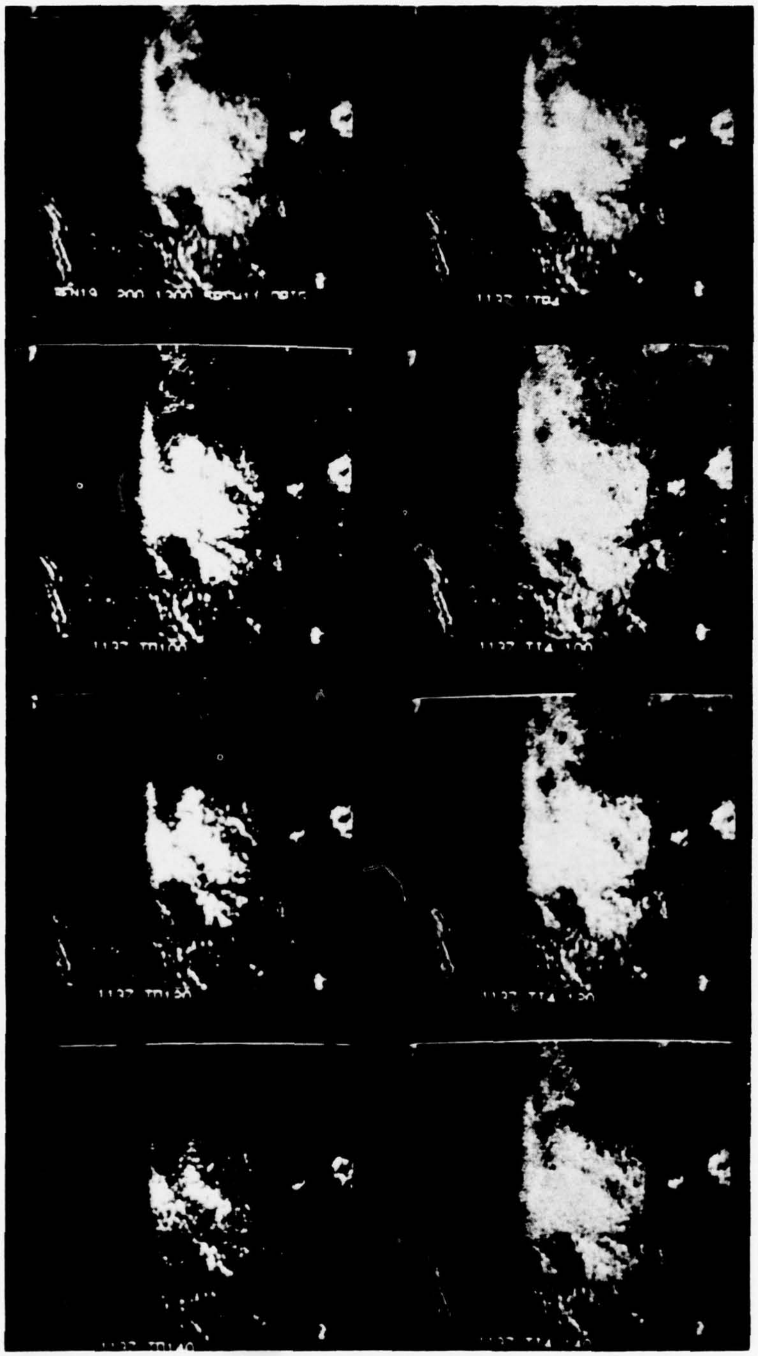


Figure 7A. Results for Case A (see text for explanation)

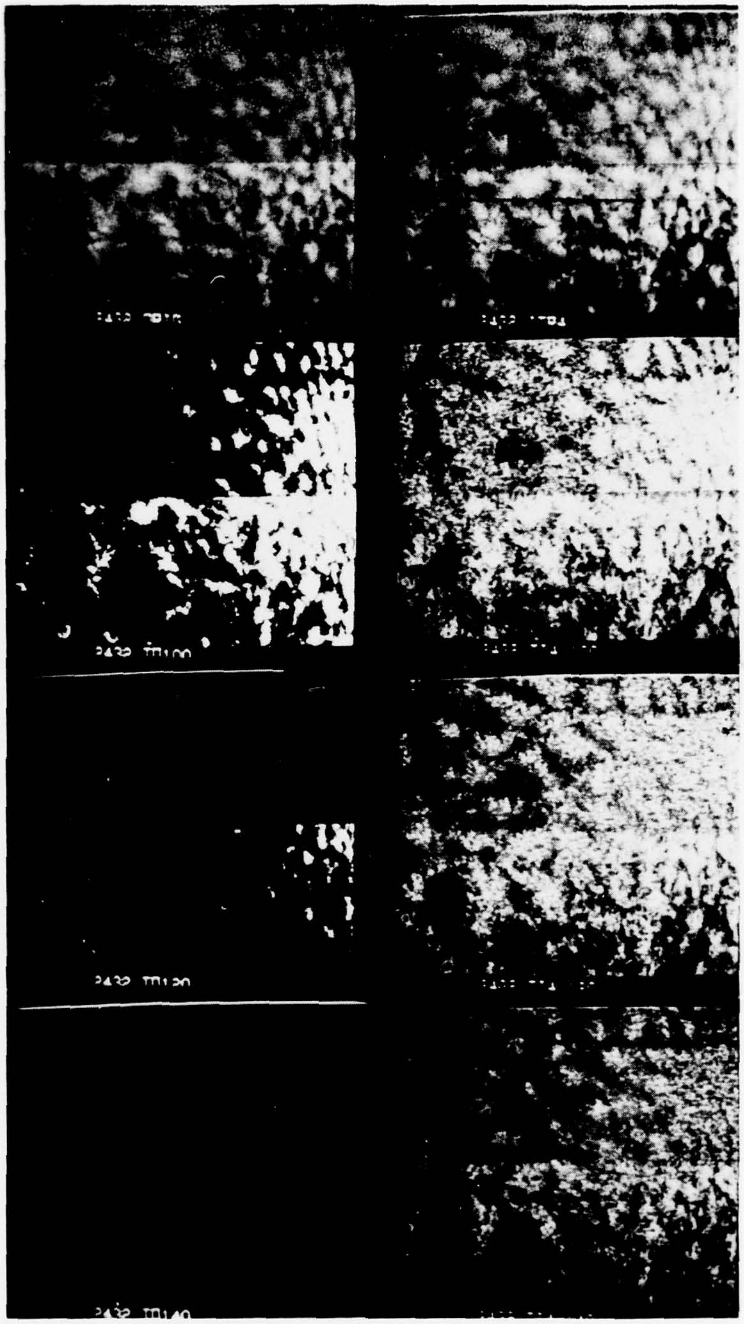


Figure 7B. Results for Case B (see text for explanation)

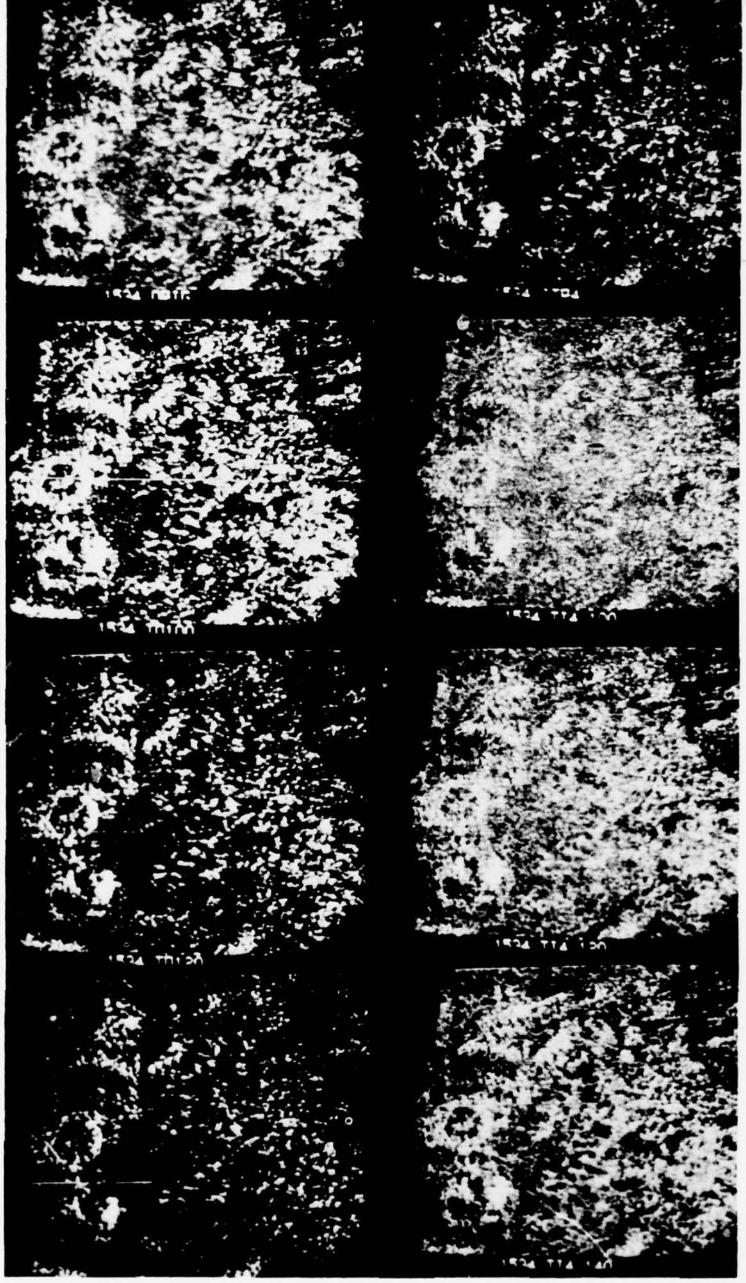


Figure 7C. Results for Case C (see text for explanation)

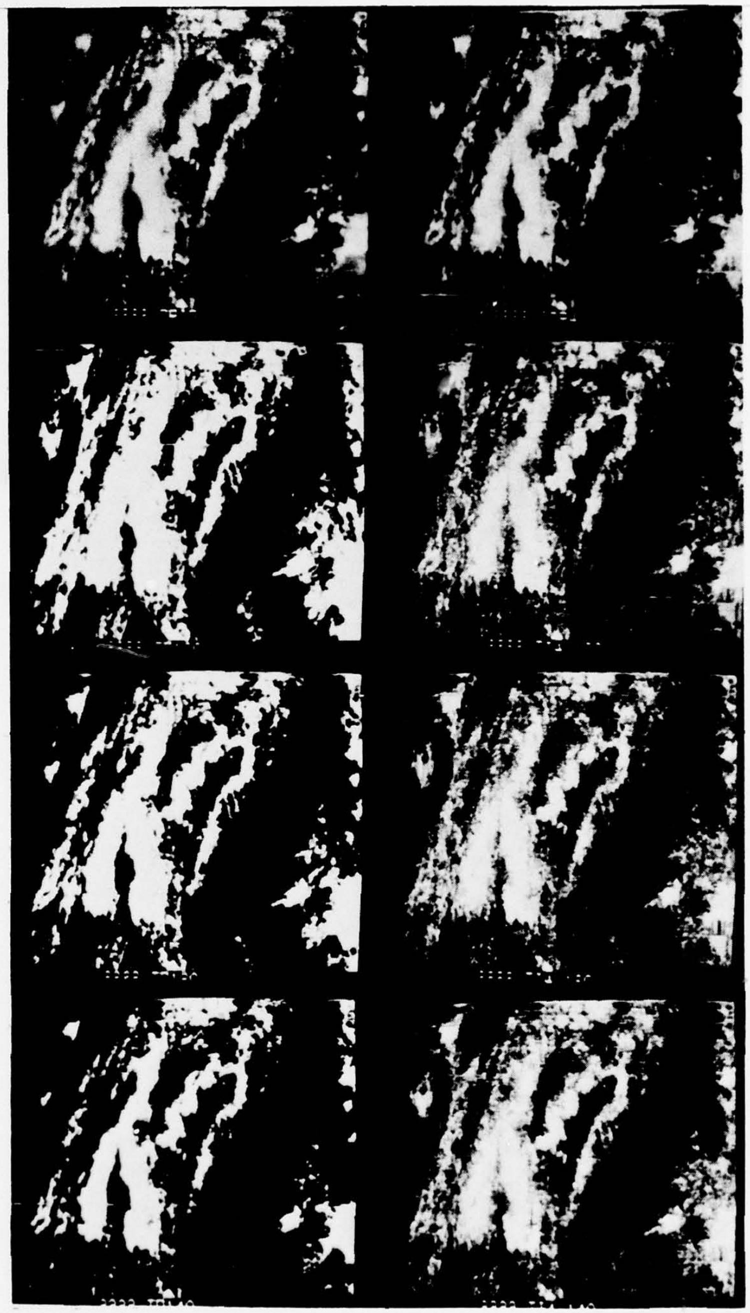


Figure 7D. Results for Case D (see text for explanation)

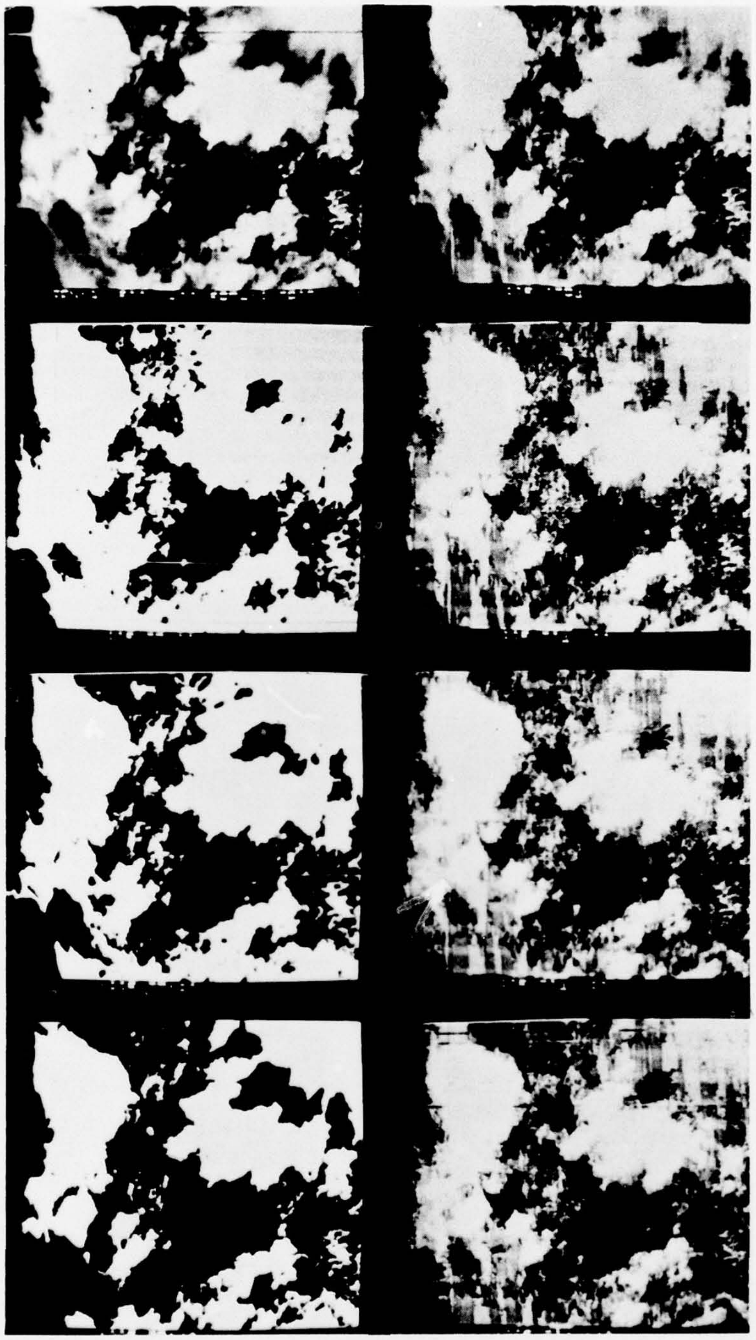


Figure 7E. Results for Case E (see text for explanation)

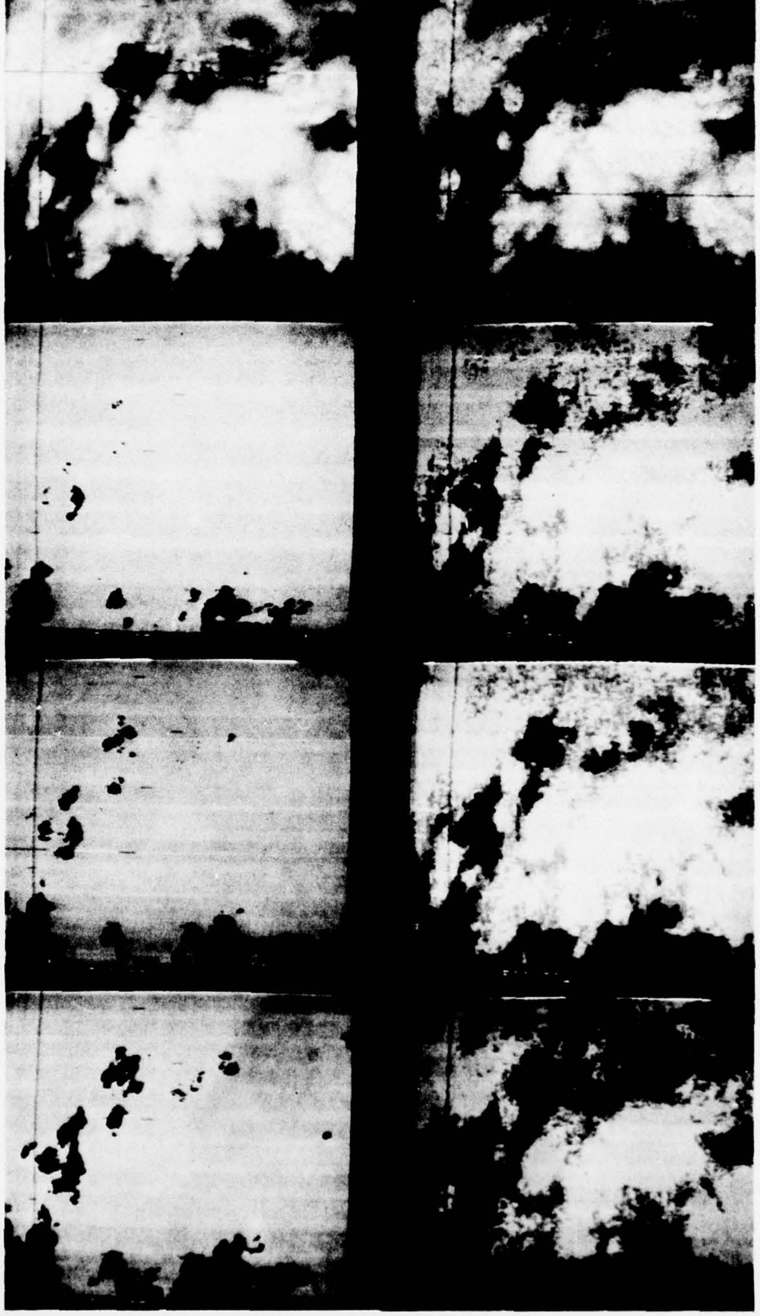


Figure 7F. Results for Case F (see text for explanation)

They are useful here since they give an indication of image brightness from comparisons of different levels of truncation. A noticeable thing about these images is that their character, in terms of both texture and tone, are greatly impaired from that of the original. There is one exception, Figure 7C, in that it preserves some resemblance as a result of very bright cumulus; but even here the change with truncation level is large. All in all, there is no level of truncation that could be selected that would produce images having a good resemblance to the original.

The bisected images, on the other hand, are excellent in representing the original six-bit images. Before evaluating them, however, consider where they came from—the fourth iteration images. These images retain very much of the integrity of the originals, and the difference between the two in some cases is hardly distinguishable. These data are still six-bit data even though the brightness values tend to be either high or low.

In Cases A and D the pair of images (original and fourth iteration) appear almost identical in all respects. Case B shows some differences in that contrasts in tone are a little sharper in the processed version; otherwise, the character of the images are very similar. Case C, the picture of clouds over land, shows a very noticeable difference in the appearance of the background brightness. All cases were processed alike in that the value for "a" was selected at 9, which is approximately the brightness level of the clear ocean. This may have contributed to the unusual appearance in the background of the fourth iteration in this case. Cases E and F show some textural irregularities which may be attributed to the display system. Even so, these processed images retain much of the essence of the originals. The main consideration, however, is how well truncated versions of the processed images (called bisected images here) measure up to the original. In particular, could one level be selected for use without making sacrifices of information content?

The bisected images, consisting of one bit per picture element, capture an amazing amount of the essence or image integrity of the original. They are not at all sensitive to levels of truncation as can be seen by comparing the three versions (truncated at levels 25, 30, and 35). They are very noticeably superior to the truncated images which give poor representations in all but select instances of truncation level and cloud type.

The importance of the bisected image lies not so much in its improvement over the truncated image, but that it captures in one bit per picture element (one-sixth the amount of data of the original) the basic characteristics of the cloud images. And in addition to this, it provides information in a simple, readily usable form having considerable reduction of image redundancies.

If it has not previously occurred to the reader, it should be pointed out that the one-bit (bisected) images are frequency modulations in the plane and provide, at the same time, both <u>spacial</u> and <u>spectral</u> properties of images! The ratio of the number

of pulses to the total number possible give brightness information which is a spacial property. The sizes of pulse/no-pulse composites provide information on basic textural or spectral aspects. The measurement in the former case is of pulses per unit length (or area) and in the latter, the reciprocal, length (or area) per unit number of pulses.

The evaluation of these images, consisting of primitives of brightness and texture, takes the form of counting; the counting of pulses or whatever is used to record the data. This is the case whether it involves the extraction of general image properties or very specific ones. And the recognition process involves the development of appropriate counting schemes for the purpose of extracting the desired information or some intermediate information which has been called image properties.

Figure 8 shows four pairs of images. They are enlargements of the original (left one of pair) and bisected image (right one) and are for 25 $\times$ 25 nmi, areas taken from Case E. They are enlargements by a factor of five on the McIDAS screen. All other images in this report have a one-to-one relationship between data and TV elements, but these have 5 $\times$ 5 McIDAS elements for each 1 $\times$ 1 data element. The bisected image for level 35 was used for these blowups. The pair of numbers underneath the image pairs are row-column references to the 1/8th mesh grid boxes in Figure 6 - Case E.

In the bisected images at this magnification, the "one-bitness" (or "two-levelness") of the images is obvious. And the variation over the surface to approximate intermediate tones occurs much as was anticipated. There is one feature that was not expected–a tendency toward arrangements into rows and columns over areas without texture such as areas of cirrus clouds (top right). Another regular pattern which has not been observed in these instances, but is virtually equivalent to alternating row and column arrangements is the checkerboard pattern. This occurrence suggests that the row/column arrangements are preferred by the calculation scheme. It is possible that this tendency is a result of applying the "cross" and "diagonal" processes with the same weight. The occurrence of long rows and columns could be useful for classification purposes. Their measurement could be used to identify, or provide some information to help identify, vaguely defined (fuzzy) entities represented in images (like cirrus) for it is under these conditions that long rows/columns occur.

As mentioned earlier, quantitative evaluations are possible in the form of rms calculations between original and one-bit images as a function of areal averages. Results of such calculations for Case E are given in Figure 9. The heavy solid lines, relatively horizontal and marked Q_{25}, Q_{30}, Q_{35}, and Q_{40} are for truncated images at levels denoted by the subscript. Iteration four is labeled IT4 and bisected images are labeled $IT4_{25}$, $IT4_{30}$, $IT4_{35}$, and $IT4_{40}$.

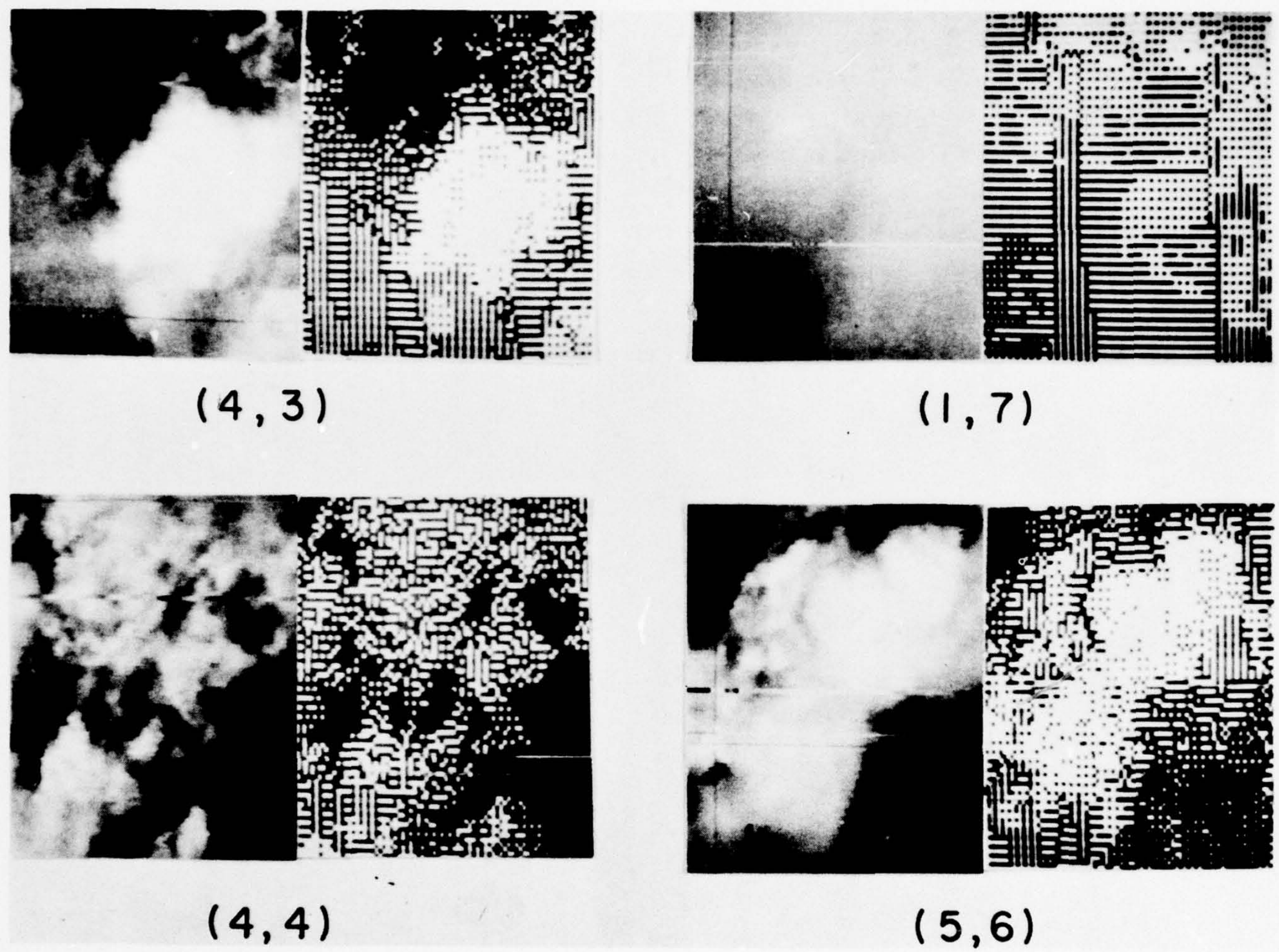


Figure 8. Enlargements of Original Image and Bisected Image of Four Square Areas (25×25 nmi) for Case E. The two numbers beneath the pairs of images are the row-column index for locating in image E of Figure 6

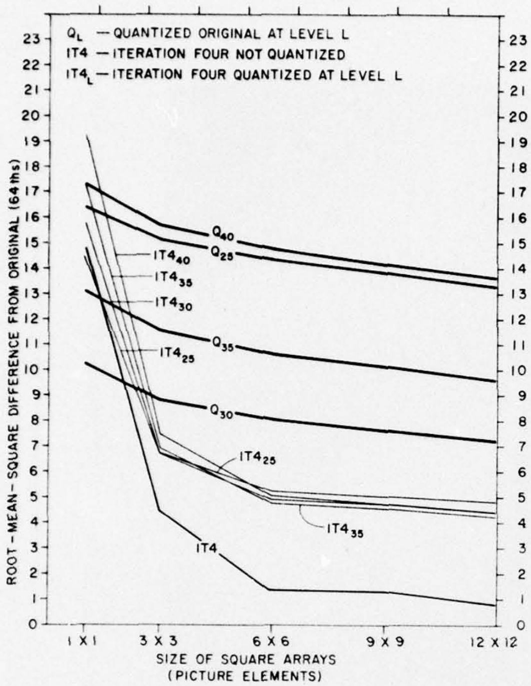


Figure 9. Quantitative Evaluations of Quality of Truncated and Bisected Images

Since the upper value, "b", does not have an absolute value in and of itself, it was obtained for these calculations in the following way. The surface brightness value was set at level 10 and the one-bit images were normalized by selecting "b" in a way to produce the same total image brightness as that of the original. All the curves of Figure 9 tend toward zero as the size of the area which is averaged increases to the total image size. The 1×1 calculations were done for all elements, then they were grouped in 3×3 arrays; after this, these were grouped to form the remaining three sets of data—6×6, 9×9, and 12×12 sets of averaged brightnesses. The ordinate of the graph is marked in brightness units based on the 64 level scale.

The curves for the truncated images, marked with a Q and subscript, are high for small areas and do not decrease much for larger areas. Iteration four, on the other hand, is high at the 1×1 area size and decreases rapidly with an increase in size of area. For 6×6 areas the rms difference is slightly less than 2. Curves for the bisected images follow the same pattern, but are not as small as those for iteration four, which is to be expected. At the 6×6 area size the value has decreased to between 5 and 6. The similarity of these four curves for the bisected images indicates a uniformity of quality independent of what level is selected for quantizing which agrees with the visual inspection. This fidelity measure points up the superiority of the bisected images over the truncated images, and provides a means for making judgments separate from individual biases or preferences.

5.5 Comments on Implementing Technique

In making the experimental calculations it was found that considerable time was required for execution of the algorithm on the CDC 6600. For operational purposes there are ways of reducing the excessive demands of time such as the use of parallel processors or hard-wired computers; however, efficient execution on conventional computers is desirable for those instances that would require their use, such as for limited production runs and experimental evaluations. Also, to execute the technique on conventional computers on a real time basis would be a very positive attribute. Such computers are more readily available than larger specialized ones and they are usually flexible as to product output. Thus, the use of a conventional computer would permit a monetary savings and could also facilitate technique improvements as system upgrading and experience demands.

The initial program execution time for an image the size of the McIDAS TV screen (672×475 picture elements), excluding identifier at bottom, was 660 seconds. Experiments* with modifications of the algorithm and optimized programming cut running time in half. It is believed another reduction in time by a factor of two is possible without seriously affecting the results. Computation time would then be 3.3 sec per 1/8th mesh box.

It is interesting to note that the algorithm can actually be processed in one sweep, that is, one iteration, providing there is an adequately large buffer system. For example, suppose there is a continuous scan of data which need not be direct from sensors but can be fitted geographically or corrected for nadir angle or other problems. All that is required is that there be sequential image scans. This can be viewed in "drum fashion" as pictured in Figure 10. Data proceeds through this system from INPUT of $G(x, y)$ which is really G as a junction of time, to the OUTPUT of $\psi(x, y)$, also a function of time. The shaded area represents a zone between both sides of the image. The "continuous" line represents scan lines. Individual data points are not shown except those which are processed at each step indicated as points within 3×3 arrays. There are four such arrays corresponding to the four "iterations". The data points are not shown to scale of course since scan lines normally contain several thousand data points. The Q in the box immediately before the OUTPUT stands for the "quantization" of the data coming into it which is "fourth iteration" data.

A visualization of this system can be had by thinking of the data stream as flowing or moving through like a rope (in steps) around a drum. The four processing arrays perform the numerical calculations of the algorithm for each step as the data

*Time and projected estimates given here are based on preliminary results obtained by Dr. Joseph Noonan and Mr. Brendan Welch of RDP, Waltham, Massachusetts, under contract to the Computer Laboratory, AFGL.

go through. Except at the two sides of the image, this process is exactly the same as that described earlier. Since flows of brightness are restricted to small areas by the numerical process, boundary irregularities would remain near the boundaries. This difference exists if no attention is paid to the gray zone as calculations are made at each step. If there were logic included to indicate when a 3×3 box is off the edge of the image and in the gray zone (and under these conditions no processing takes place for that box) then the results of this execution of the algorithm in "one iteration" would be exactly the same as those of the four iteration version.

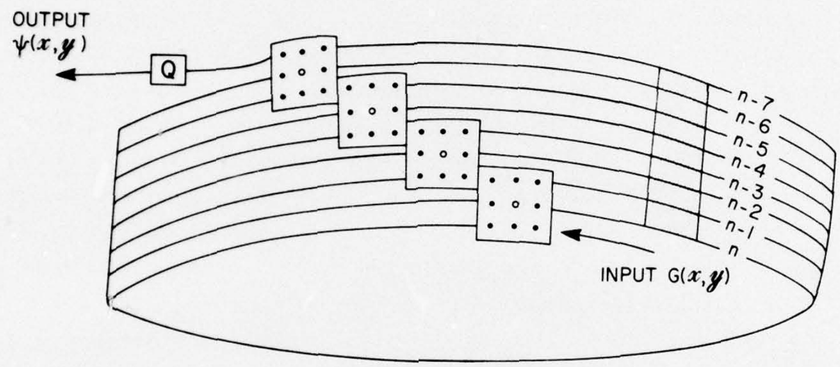


Figure 10. Diagram of Computational Set-up for Sequential Processing of Images on Scan-Line Basis

This system requires an accessible storage buffer for eight scan lines plus twelve data points. There is some chance that the four passes of the algorithm could be reduced to three or even two. This would reduce the buffer requirement. On the other hand, if the background brightness were included as a variable, a duplicate buffer system would be needed in synchronization with this one.

There are several possibilities for implementing the "bisecting" technique, but initial experience indicates that it will present a challenge to obtain a means for a real-time analysis of satellite imagery on the conventional computer. The other possibilities are to use signal processing equipment, specially designed hardware or array (parallel) processing computers which have developed rapidly in the last few years for problems of this kind. It is too early to conjecture as to what can be done in terms of implementation of the technique in these areas.

6. CONCLUDING COMMENTS

Little progress has been made on the problem of redundancy in satellite images since Glaser[1] first discussed it. It remains a major obstacle to automated analysis. Although the science and art of image analysis have been moving forward at a rapid rate, and some of this technology is being absorbed into satellite image analysis, techniques from Communication Theory for reducing redundancy have not been satisfactory for one reason or another. Relatively little effort in this area has been made within the field of satellite meteorology itself.

The problem has largely been ignored or otherwise left to the "bigger and faster" computers of the future. If the past several years are any indication of the future, this may well represent a false hope. Although computer technology has increased significantly in recent years, the improvement in automated satellite image analysis has been minimal. Even though larger, faster computers are desirable and perhaps even necessary, they will only contribute partially to the ultimate solution. The complexities of imagery far exceed the capabilities of solutions based on hardware and the "method of exhaustion". That is, the reduction of the problem to direct brute force methods of connecting multiple channel data with desired results does not appear to be the most feasible answer.

In this report a part of the redundancy problem described by Glaser has been examined. It may be characterized as the "local redundancy" part to distinguish it from redundancies occurring over large regions in images which could be termed "global redundancy". The approach given here of removing "local redundancies" without consideration of "global redundancies" appears to be desirable from many points of view, especially in considering automated extraction of information.

The development of extraction techniques using bisected images is beyond the scope of this report. It is, however, a prime area for needed development to bring closer to realization the system schematically shown in Figure 1. It is instructive to review that Figure in light of the results obtained since it was introduced. It is a very general plan for image analysis. The individual boxes representing operations on the data could consist of a number of different techniques. The parts which perform operations on two dimensional arrays, although labeled simply in the diagram, may in fact consist of several parts having different purposes. As for physically implementing this system, the greatest problems of analysis occurs in the early stages (extreme left of Figure 1).

Data amounts are so high in the initial stages of satellite image analysis that little in the way of processing can be done without considerable simplification. One such simplification that is sometimes made is spacial averaging. This method is efficient in reducing the amount of data that has to be handled, but it defeats the original intention of obtaining high resolution data. Another procedure, even less

sophisticated, is to delete specific picture elements and lines, typically every other picture element and every other line. This procedure sacrifices spacial resolution as well. The data rate in this case is reduced to one-fourth the original.

The technique described in this report is a simplification that retains many desirable features of conventional imagery and circumvents difficulties encountered in standard methods of analysis. It is of a general nature and applicable to a wide range of analysis problems, but since we have little experience with the technique, it is not possible to make any definite assertions about results to expect for specific applications. The technique should receive increased interest, experimentation, and utilization by many disciplines to better gauge its usefulness. Such tests would probably consist mainly of numerical experiments that would include experiments with different kinds of data under different model configurations.

Closely related to this work is the study of "means" of solutions. For instance, what equipment is needed to obtain solutions in various cases? Experience is lacking here also.

Separate from these areas is another that requires development if a solid application of the technique is to be made to automated analysis. Techniques for "recognition" or "classification" will be needed. Specific experience is lacking in this area but there are techniques in data analysis that may be directly applicable.

There are numerous possibilities for the analysis (classification, evaluation, interpretation, extraction) of bisected images. In briefly considering this subject it is instructive to refer again to Figure 1. The ultimate object is to reduce satellite images to useful meteorological information by taking into account information derived from a number of channels. Image properties discussed earlier and diagrammed in Figure 1 can be definitions, so to speak, of cloud appearance from a satellite point of view. There are no established principals here and further development is needed.

More experiments and specific applications on the obtaining and use of bisected images will provide invaluable needed information. Theory and image statistics are not well enough developed at this time to be of much use. Criteria for judging the "goodness" of results and making evaluations of methods of analysis which are established on the basis of experience are areas where theory will eventually be of much value.

The ideas presented in this report and the technique for bisecting an image (or compressing picture element bits down to one) are very general concepts and can be used in a wide variety of situations. There will undoubtedly be limits that will be learned from experience instead of from theory applied to image statistics. Applications as well as results obtained from experience will provide the best guide for the development and use of bisected images.

There are several areas in satellite image analysis that the bisected image technique is potentially useful. Some of them are:

(a) Special purpose image channels that are restricted to one bit per picture element,

(b) Graphics display systems having a limitation on the handling capacity and on the number of grey shades available,

(c) Cloud classification and information extraction for 3-D Nephanalysis,

(d) Non-operational analysis.

The comments above about a need for experience apply in each of these cases.

In the past two decades much effort has been expended in the "interpretation" of satellite images for purposes of meteorological "analysis and forecasting". Much of this work has relied on human judgment. Consequently there is meteorological information based on experience and in many cases sound principles that is available to the image analyst but physically impossible to pass on to users.

A mass of results from the use of satellite data has been obtained since the TIROS years of the early sixties. Most of these results have not found their way to the decision maker in the field either directly in terms of on the spot interpretations or indirectly in the form of improved objective analyses. This represents a strong point in favor of turning more to objective techniques in analyzing satellite data.

Subjective human analyses, however, have certain attributes too. An experienced analyst in satellite photo interpretation can take a glance at a few pictures of some area of the globe and make some summarizing statements packed with meteorological information. That is, considerable uncertainty of the meteorological situation for the area in question is removed from the mind of another trained person who hears the comments but does not see the pictures. To date, there is no objective technique available that can provide such information so quickly or concisely.

Most scientists and technicians currently believe a compromise between the objective and subjective analyses methods will be necessary for some time, that is, automated methods should be used as much as possible, however, the human analyst is necessary for certain situations and events.

This guessing about future developments and uses of the procedure can only be given little weight. Experience which is sure to come as various experiments are performed will undoubtedly reveal a short-sightedness in these projections and will show a need for their replacement.

The potential value of this technique is great. Since so much rests on its actual implementation an accelerated program appears justified to obtain experience and gain confidence in its strengths and on understanding of its limitations. Such work is necessary before any significant operational experiments can be designed with confidence.

References

1. Glaser, Arnold H. (1957) Meteorological Utilization of Images of the Earth's Surface Transmitted from a Satellite Vehicle, Harvard University, Blue Hill Observatory, 145 pp.

2. Marggraf, W. A. (1967) Information Content, Elemental Feature Extraction and Coding of Meteorological Satellite Television Data, General Dynamics Report No. GD/C-ERR-AN-1053, unpaged.

3. Kutz, R. L., Sciulli, J. A., and Stampfl, R. A. (1968) Adaptive data compression for video signals, Advances in Communication Systems, Vol. 3, edited by A. V. Balakrishnan, Academic Press, New York, pp 29-66.

4. Häberle, H., Ulrich, P. C., and Zachunke, W. (1974) Digital TV transmission via satellites, Electrical Communication, 49, No. 3, International Telephone and Telegraph Company, pp 326-331.

5. Musmann, H. G. (1973) Theoretical aspects of intraframe coding, Deutsche Luft- und Raumfahrt, Forschungsbericht Munchen, Fentralstelle fur Luft-fahrtdokumentation und -information.

6. Kummerow, T. (1972) Statistics for efficient linear and non-linear picture encoding, Proceedings of the International Telemetering Conference, 10-12 October 1972, 8:149-161, International Foundation for Telemetering, Woodland Hills, California.

7. Pratt, W. K. (1960) A comparison of digital image transforms, Proc. Mervin J. Kelly Commun. Conf., 1970, pp 17.4.1-17.4.5.

8. Habibi, Ali (1971) Image coding by linear transformation and block quantization, IEEE Trans. Commun. Tech. Com-19(1):50-62.

9. Davisson, L. D., and Gray, R. M. (1976) Data Compression, Benchmark Papers in Electrical Engineering and Computer Service/14, Dowden, Hutchinson and Ross, Stroudsburg, Pennsylvania.

10. Haralick, R. M., and Shanmugan, K. W. (1974) Combined spectral and spacial processing of ERTS imagery data, Remote Sensing of the Environment 3:3-13.

11. Haralick, R. M. (1973) Glossary and Index to Remotely Sensed Image Pattern Recognition Concepts, Pattern Recognition, Vol. 5, Pergamon Press, pp 391-403.

12. Rosenfeld, Azriel, and Kak, A. C. (1976) Digital Picture Processing, Academic Press, New York.

13. Carne, E. B. (1975) Artificial Intelligence Techniques, MacMillan and Co., London.

14. Banerji, R. B. (1969) Theory of Problem Solving, Elsevier Pub. Co., New York.

15. Mendel, J. M., and Fu, K. S. (Eds.) (1970) Adaptive, Learning and Pattern Recognition Systems, Academic Press, New York.

16. Nilsson, N. J. (1971) Problem-Solving Methods in Artificial Intelligence, McGraw Hill, New York.

17. Sampson, J. R. (1976) Adaptive-Information Processing, An Introductory Survey, Springer-Verlag, New York.

18. Tsypkin, Ya. Z. (1971) Adaptation and Learning in Automatic Systems, Translation of Adaptatsia i obuchenie v avtomaticheskikh sistemakh Nauka, Moskow, 1968, Academic Press, New York.

19. Pickett, R. M., and Blackman, E. S. (1976) Automated Processing of Satellite Imagery Data at Air Force Global Weather Central (AFGWC): Survey, Recommendations and R&D Design Evaluation Report, 23 April 1976, Bolt Berenak and Newman Report 3275, 62 pp.

20. Coburn, A. R. (1971) Improved Three Dimensional Nephanalysis Model, Air Force Global Weather Central Publication, AFGWC TM-71-2, 72 pp

21. Canipe, Yates J. (1976) A real time satellite processor, Seventh Conference on Aerospace and Aeronautical Meteorology and Symposium on Remote Sensing from Satellites, American Meteorological Society, 16-19 November 1976, pp 298-301.

22. Janant, Nuggehally S., Ed., (1976) Waveform Quantization and Coding, IEEE Press, New York.

23. Abramson, Norman (1963) Information Theory and Coding, McGraw-Hill, New York.

24. Young, J. F. (1971) Information Theory, Wiley Interscience, New York.

25. Aczel, J., and Daroczy, Z. (1975) On Measures of Information and their Characteristics, Academic Press, New York.

26. Bendig, A. W. (1953) Twenty questions: on information analysis, J. Ex. Psy. 46(No. 5):345-348.

27. Shannon, C. E. (1948) A mathematical theory of communication, Bell System Tech. Journal, 27:379-423 and 623-656.

28. Shannon, C. E., and Weaver, W. (1949) The Mathematical Theory of Communication, Univ. of Illinois Press, Urbana, Illinois.

29. Khinchin, A. I. (1957) Mathematical Foundations of Information Theory, Dover Publications, New York.

30. Feinstein, A. (1958) Foundations of Information Theory, McGraw-Hill, New York.

31. Blasbalg, H., and Van Blerkom, R. (1962) Message compression, IRE Trans. Space Electron. Telemetry, 8:228-238.

32. Gray, R. M., and Davisson, L. D. (1974) A mathematical theory of data compression, Proc. 1974 Intern. Conf. Commun. 1974, pp 40A-1-40A-5.

33. Berger, Toby (1971) *Rate Distortion Theory, A Mathematical Basis for Data Compression,* Prentice-Hall, New Jersey.

34. Conover, J. H. (1962) Cloud Interpretation from Satellite Altitudes, CR Research Note 81, AFCRL, 77 pp; and Supplement 1, 1963, 19 pp.

35. Van Soest, J. L. (1956) Some consequences of the finiteness of information, Information Theory, edited by Colin Cherry, Butterworth Scientific Publications, London, pp 3-7.

36. Polya, George (1945) How to Solve it, Princeton University Press, Princeton, New Jersey.

37. Polya, George (1954) Mathematics and Plausible Reasoning; Vol. 1, Induction and Analogy in Mathematics; Vol. 2, Patterns of Plausible Interference, Princeton University Press, Princeton, New Jersey.

38. Polya, George (1962) Mathematical Discovery, 2 vols., Vol. 2 copyright 1965, Wiley and Sons, New York.

39. Koopman, Bernard O. (1977) Intuition in mathematical operations research, Operations Research, **25**(No. 2):189-206.